Experiences with a Multi-Protocol network monitor

Andrew Moore
Intel Research Fellow
Computer Laboratory
University of Cambridge

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

In collaboration with:
Ian Pratt, Jon Crowcroft, James Hall, Tim Granger, Derek McAuley, Dina Papagiannaki, among others.

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Contents

- one-slider on Projects @ Cambridge
- Nprobe/GRIDprobe Monitor
- Experiences with...
  - Behaviour example
  - Content example
  - Visualizer example
- Where next...

Projects at Cambridge on Monitoring

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

- Nprobe/GRIDprobe/Xenoprobe
  - Computer Laboratory
  - (Collectively known as *probe or (star)probe)
- CoMo
  - Intel Research Cambridge + friends

Planning for convergence, also taking in Hyperion (U.Mass)

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

GRIDprobe Objective

- (Nprobe prototyped/grew-into GRIDprobe)
- Scalable Monitoring Architecture (tool building)
  - 1 & 10Gbps and viable strategy for 40Gbps
- Multi-protocol monitoring
  - Understand network and application behaviour

At the same time.

- Originally University of Cambridge & Marconi (RIP)
- Now Cambridge with association from Intel
- Duration Oct. 2002 – Sep. 2005

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Status

- Several working test deployments (1Gbps)
- Prototype for 10 Gbps
- Code base is planned for a public release
- Experience with the dataset/database/dataware-house issues
- Adding new protocol modules
- Using Experience to drive next architecture

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Where does this tool fit in?

"When you have a hammer, every problem looks like a nail."

- We want current network data
  - High-resolution timer
  - High-speed (current deployment: 1 Gbps)
- We want to collect enough information to see the interaction between layers
- We want to use commodity (no custom) hardware to maximize deployment and minimize cost

Nprobe: our current implementation

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

- Current Nprobe system performs full line-rate capture on commodity hardware
- Nprobe is a multi-protocol monitor: collecting network, transport & application data
- Nprobe processes network, transport & application layers to provide compression as well as extracting useful information (e.g., application features)

What we are NOT

- We are not just some IDS – they do a few things that superficially look the same – ultimately these things are not the same.
- We want to collect as much as possible – they want to collect the minimum and to compare as quick as possible.
- we want to interpret the full application – they want to string-match then move on.

What is the problem?

In a perfect world:

Cheaply (using commodity PCs)

Record 1, 10, (MAXINT) Gbps

Full duplex

Onto disk

With minimal loss

Ouch!

Not as bad as all that: its not a perfect world

How do we do it?

"Discard is the most effective compression."

Be selective (for an http example)

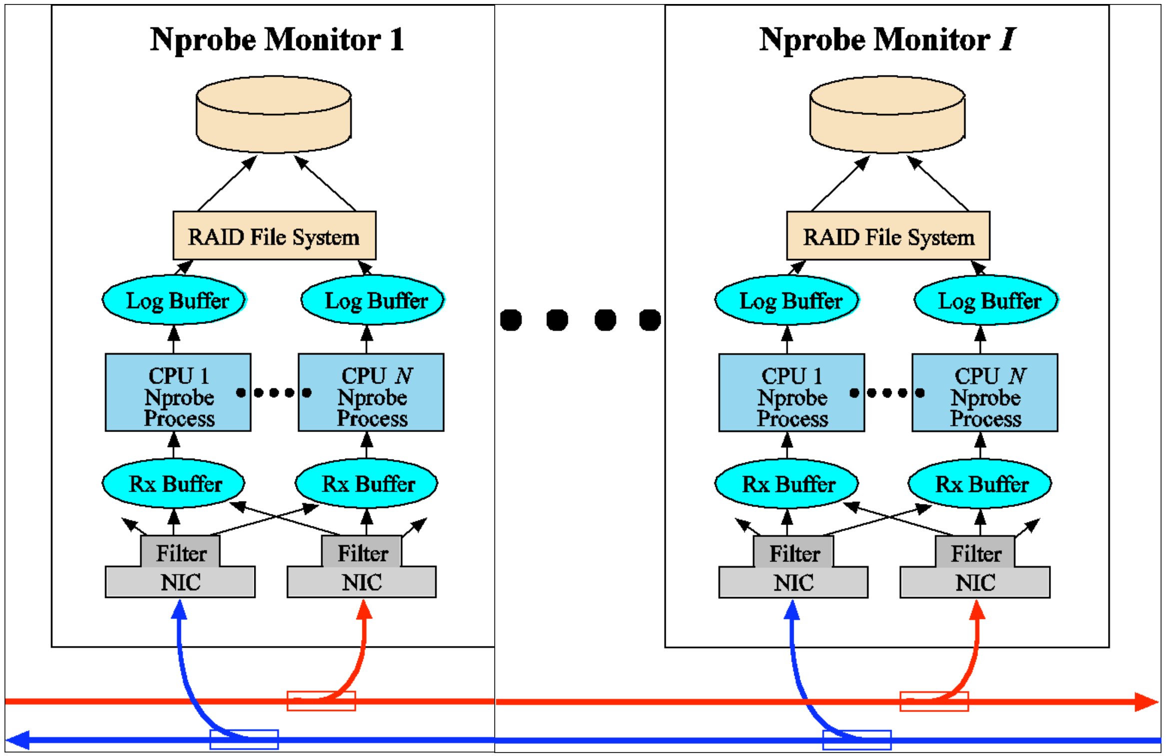
1. Remove redundant header information
2. Temporally compress header information
3. Extract http transactions from data stream
4. Remove (or summarise) uninteresting information (consider the use)

How else?

"A problem shared is a problem halved."

Split the workload

- among CPUs
- among machines



Problems?

- Complex filter design made easier using ongoing measurements

Limitations?

Limitations Abilities

- Host-host data must be less than or equal to the capacity of a single monitor (CPU)
- No monitoring DataTAG 10 Gbps host-host experiments
- For ISP and dial-up or cable modem last-miles, as well as with (UK) academics with 100 Mbps to the desktop, this approach works
- Target deployment has nx10,000 of flows and the monitor is close to the server or close to the client (on access/choke-points).

Example 1: Modelling TCP Connections

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

- Dynamic model of TCP connection activity
  - Input from probe-collected data
    - Packet timings
    - Packet header data
    - Higher level protocol activity
  - Output — identified, differentiated and quantified
    - Network times
    - TCP Artefacts
    - Application delays

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Causative Associations

- Probe sees TCP packets traveling to host and those returning
  - Arriving packets
    - Modify host TCP state
    - Cause work to be done
    - Trigger transmissions — causative associations
    - Drive model
  - Departing packets
    - Verify/modify model
    - Are arrivals at peer

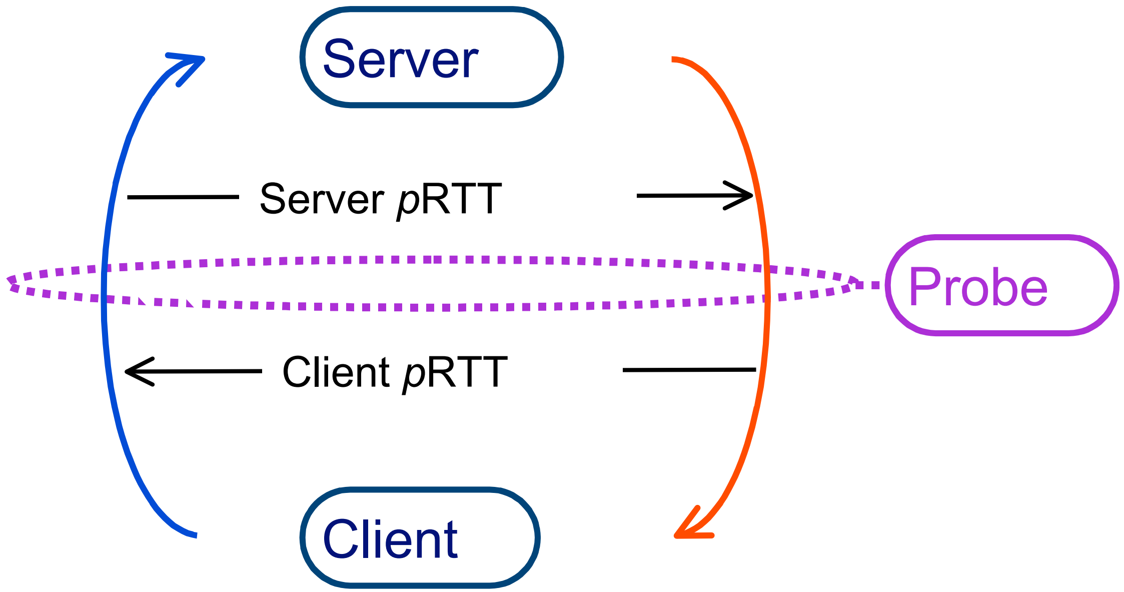
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Example 1:

- ACK packet arrives during slow start
  - sender's congestion window expands
  - releases flight of data segment packet(s)
- Data segment N arrives {N mod 2 = 0}
  - ACK released
- Data segment N transmitted
  - data segment N+1 released
- HTTP request arrives
  - First packet of response released (after delay)

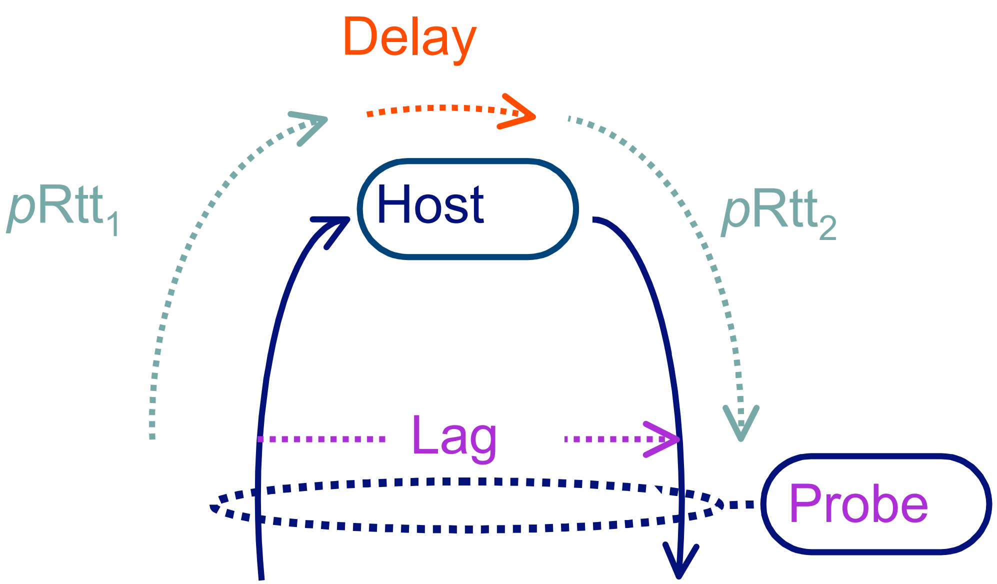
Partial Round Trip Times

- Probe can be anywhere
  - Hence deal in *p*RTTs)
  - Can glue them together



Lags, Delays and *p*RTTs

- For causative associations
  - lag = *p*RTT + delay
  - If no delay:
    - *p*RTT = lag
  - If delay:
    - interpolate *p*RTT
    - Calculate delay
  - Model informs



*p*RTT Drawbacks/Restrictions

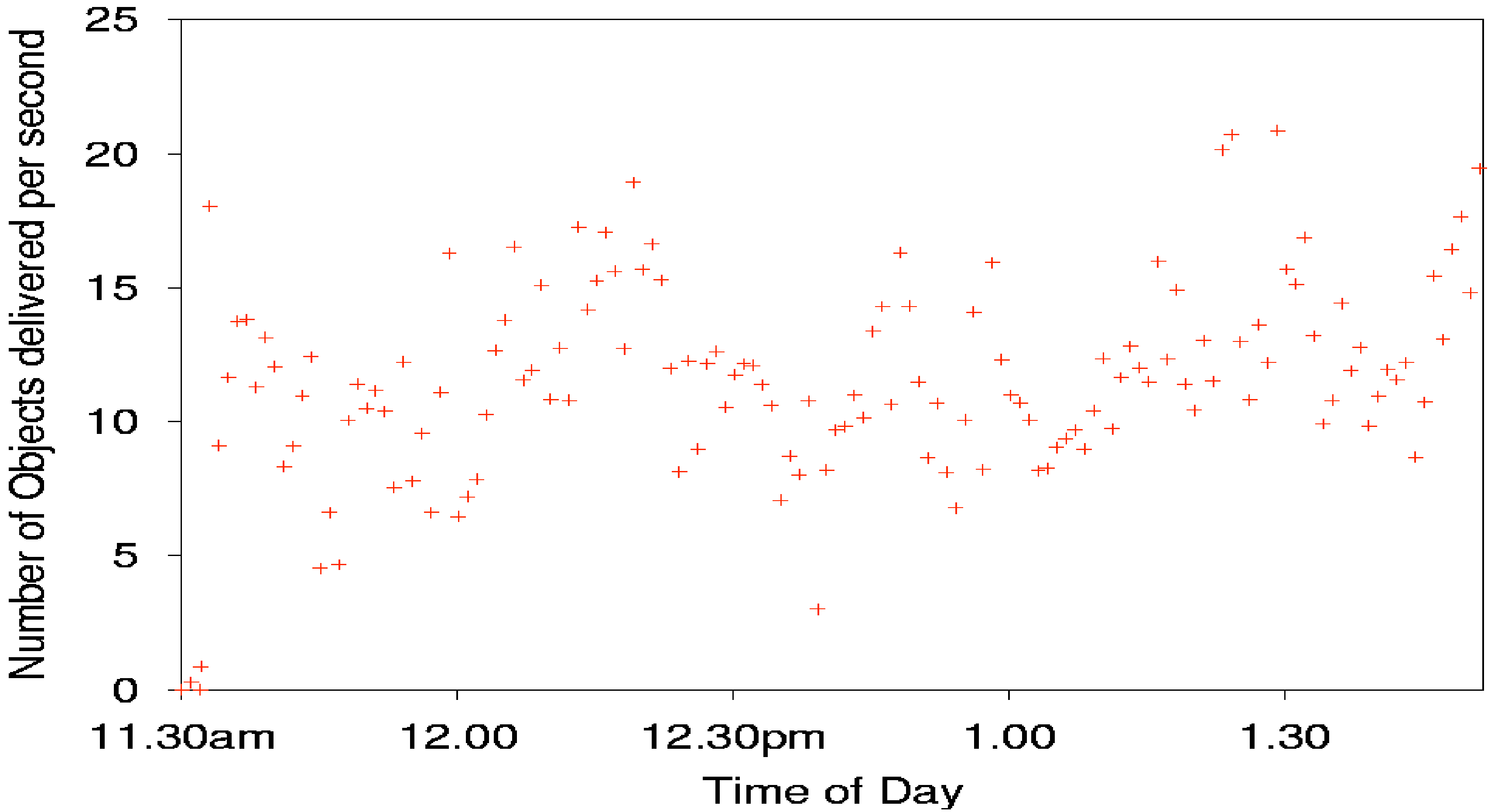
- Only works in slow-start, thus relies on longer data flows
- relies on implementation “inside knowledge” fortunately only a few implementations (BSD derivative, Linux derivative, Microsoft derivative)

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

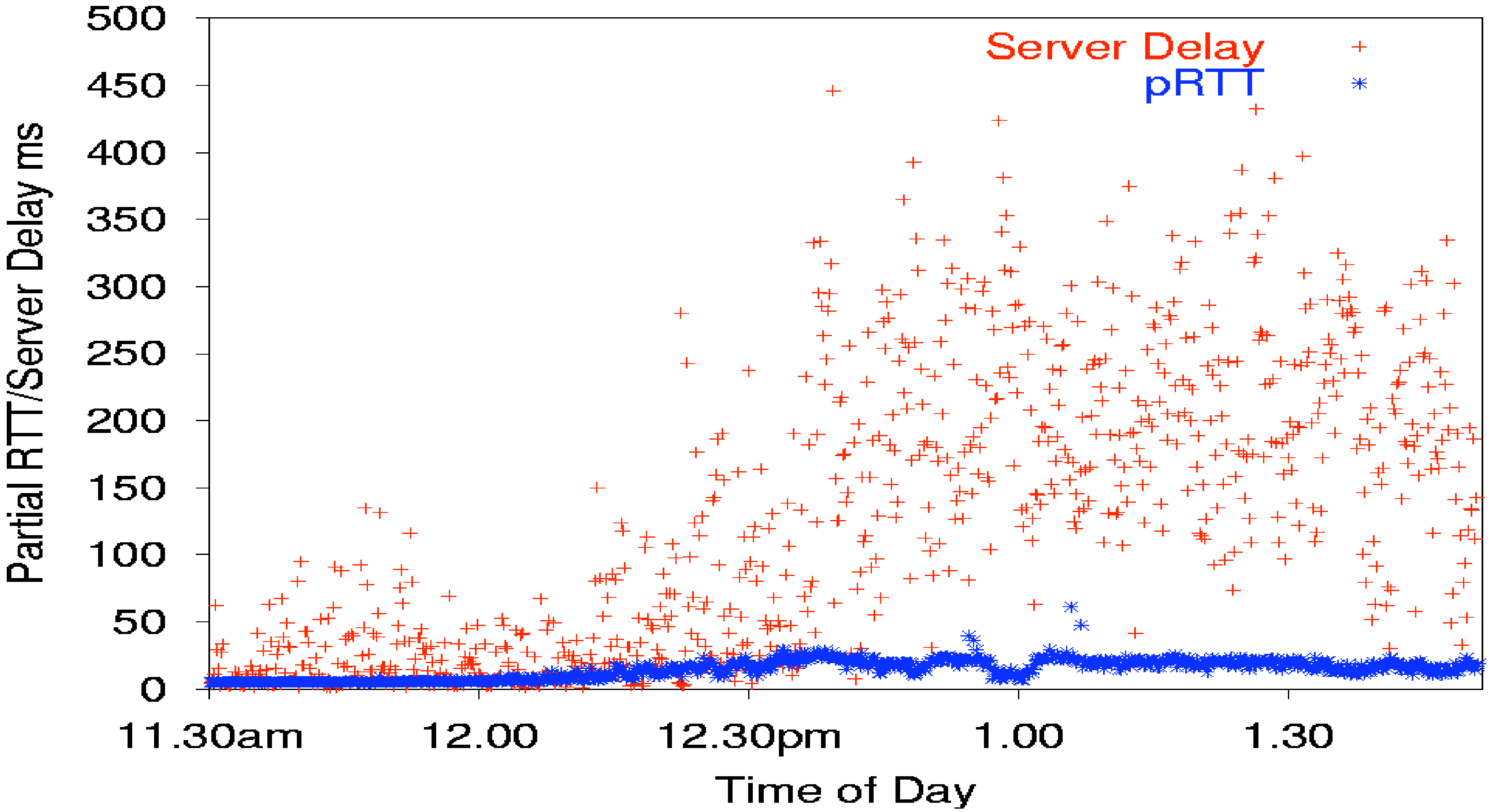
Results — Live Traffic

- All HTTP traffic to BBC news server from University site
  - 24 Hour trace
  - Results for period 1130 – 1350
    - Expect load increase as users browse during lunch break
- Independent of local load
- Look at SYN re-transmissions

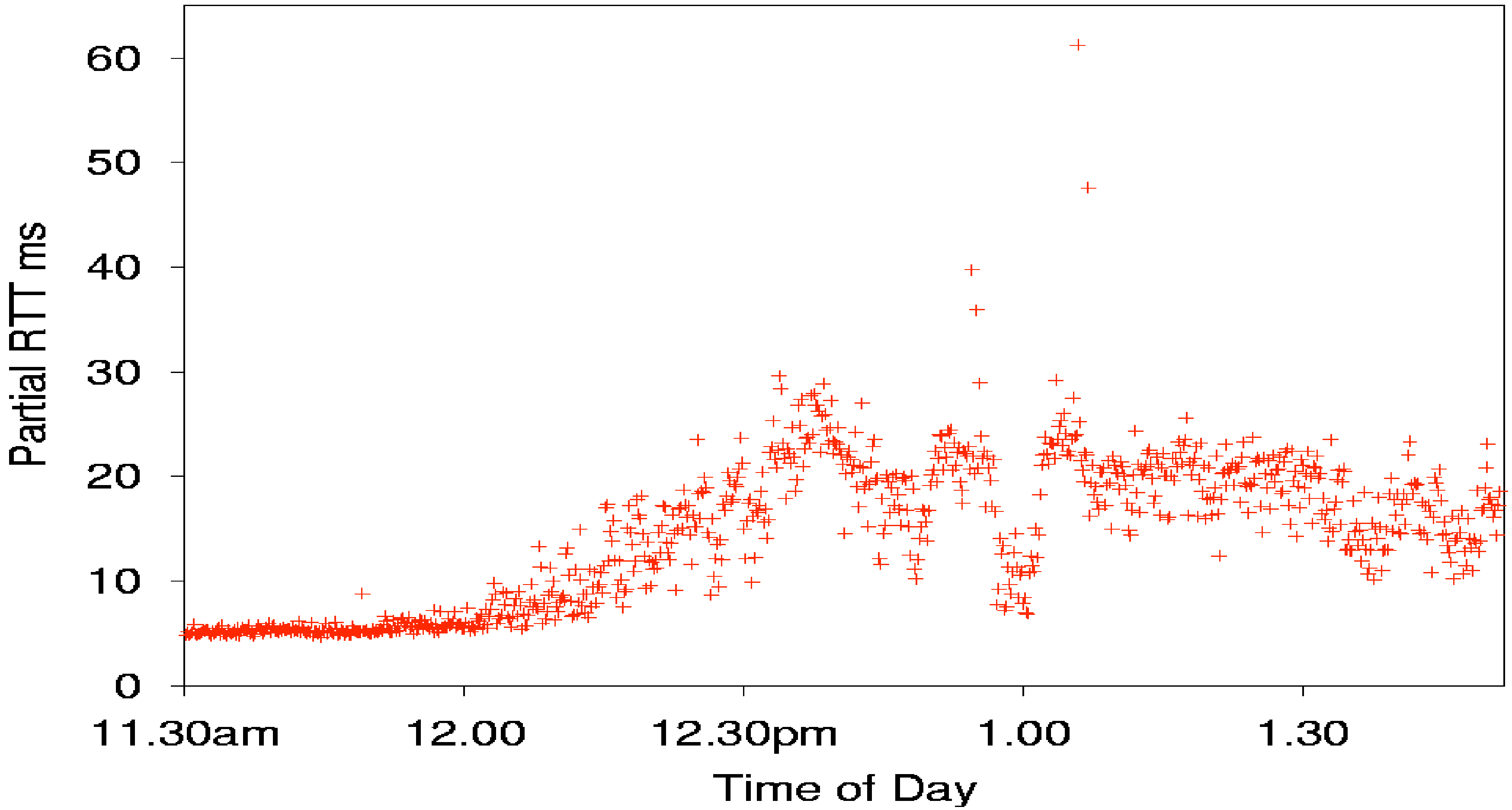
Local Load — Live Traffic



Server Delays and *p*RTTs

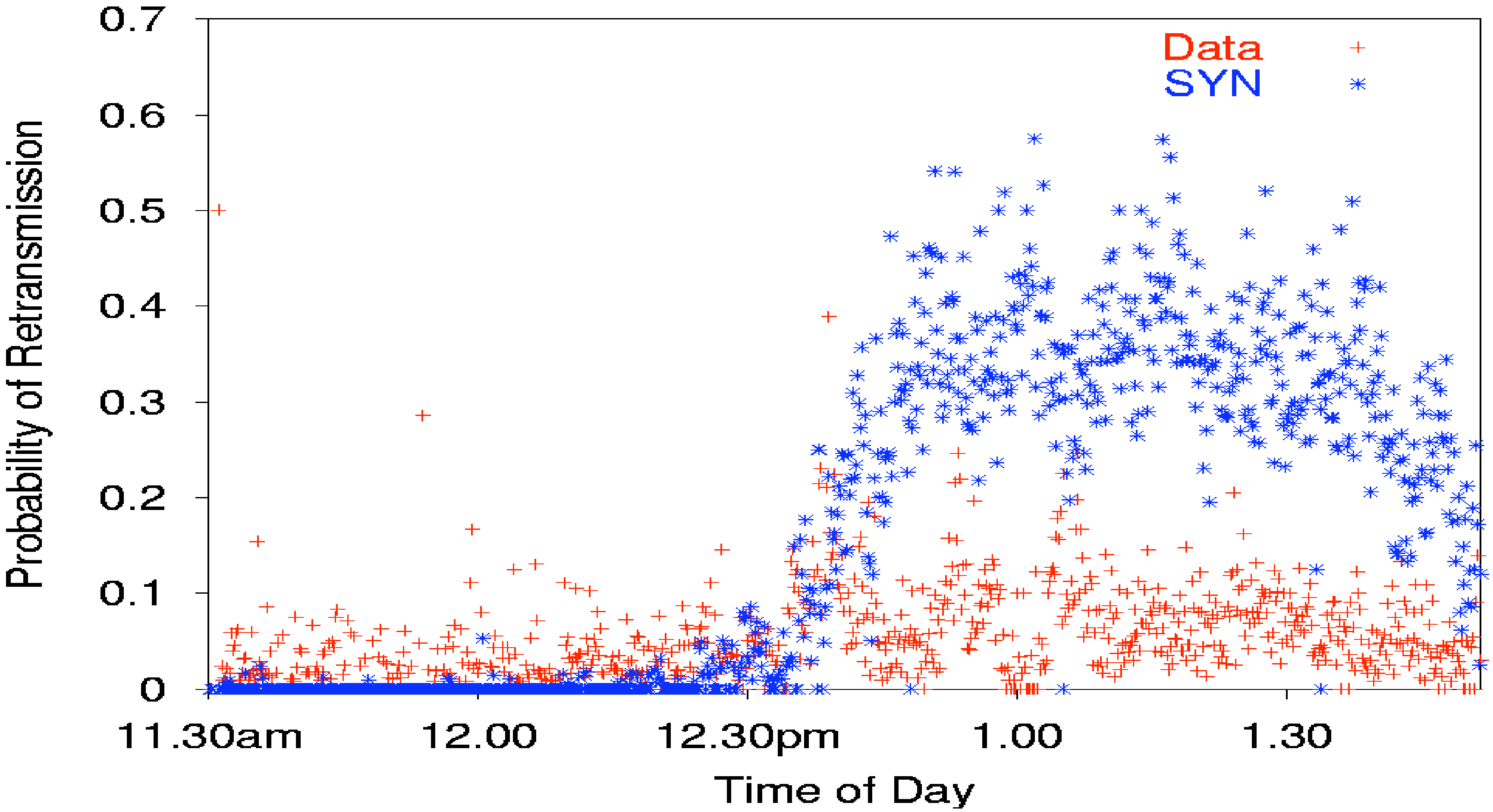


Server *p*RTTs — Live Traffic



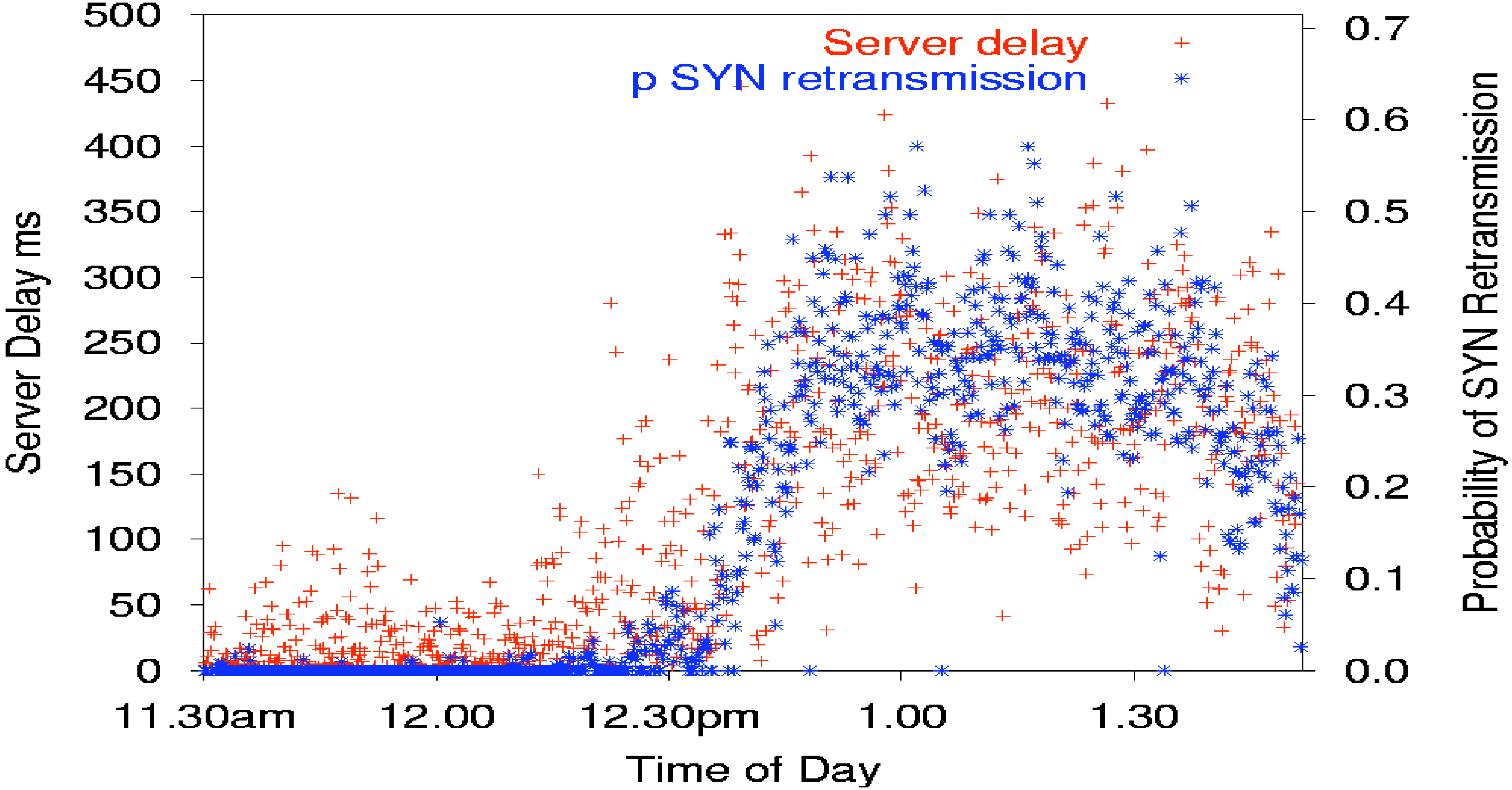
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Probabiliy-of-Retransmission



Server Delay and p-SYN Retransmission

UNIVERSITY OF CAMBRIDGE
Computer Laboratory



Behaviour Summary

- By observing a combination of TCP and HTTP protocols simultaneously, we:
  - determine the type of load-shedding this website uses.
  - understand diminished performance in the face of no local network effects.
  - Draw conclusions on the impact this approach (to load-shedding) has on persistent vs. non-persistent (compared with a nominal 25%, this site had less than 5% persistent)

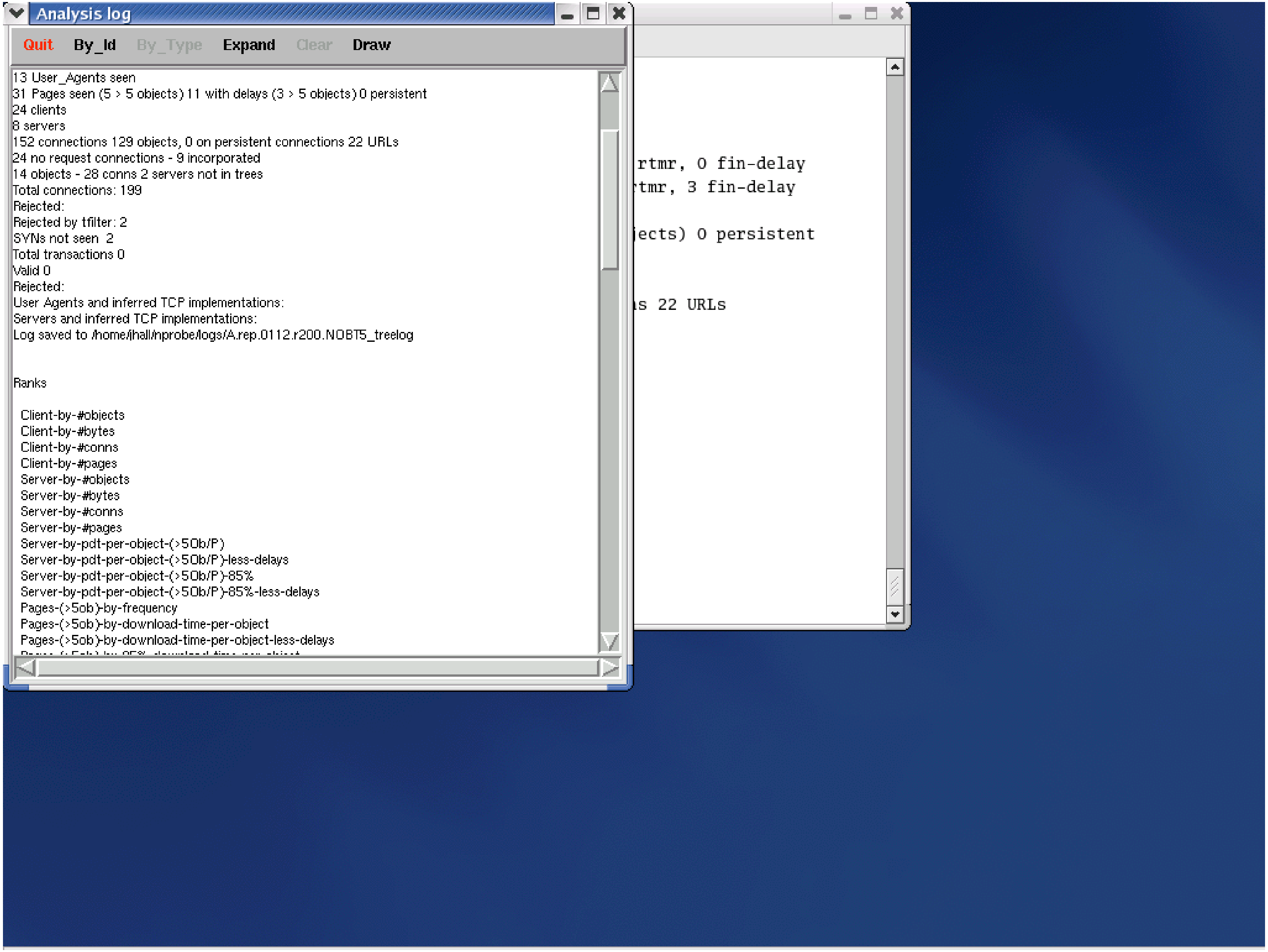
GRIDprobe Visualization Tools

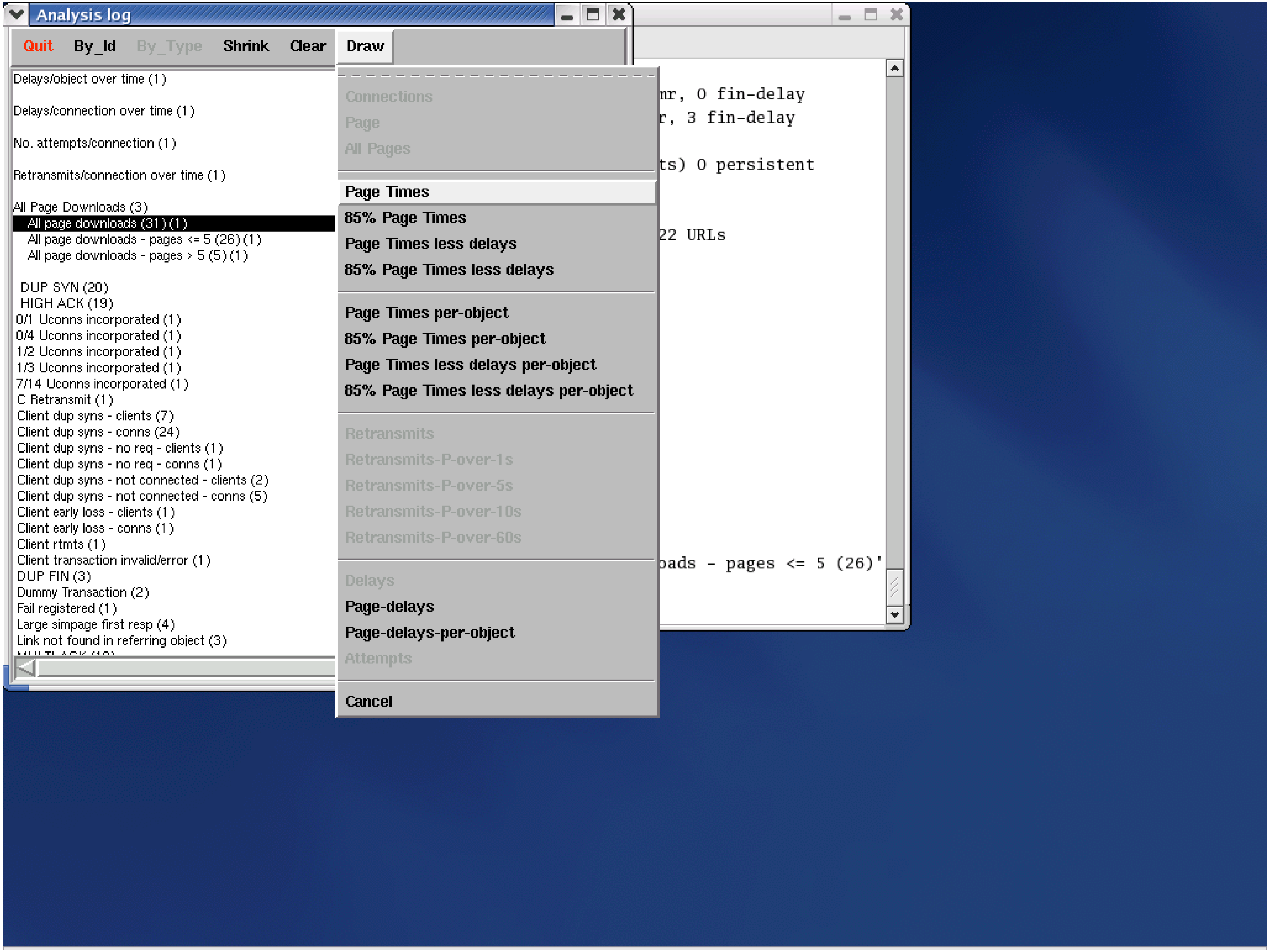
- Reads the stored format – the stored format is already partially processed
- Extracts features of interest (timing, packet headers,...)
- Constructs relationship trees for (web) pages
- provides:
  - interactive data plotter (ala gnuplot++)
  - tcp connection plotter
  - web transaction plotter

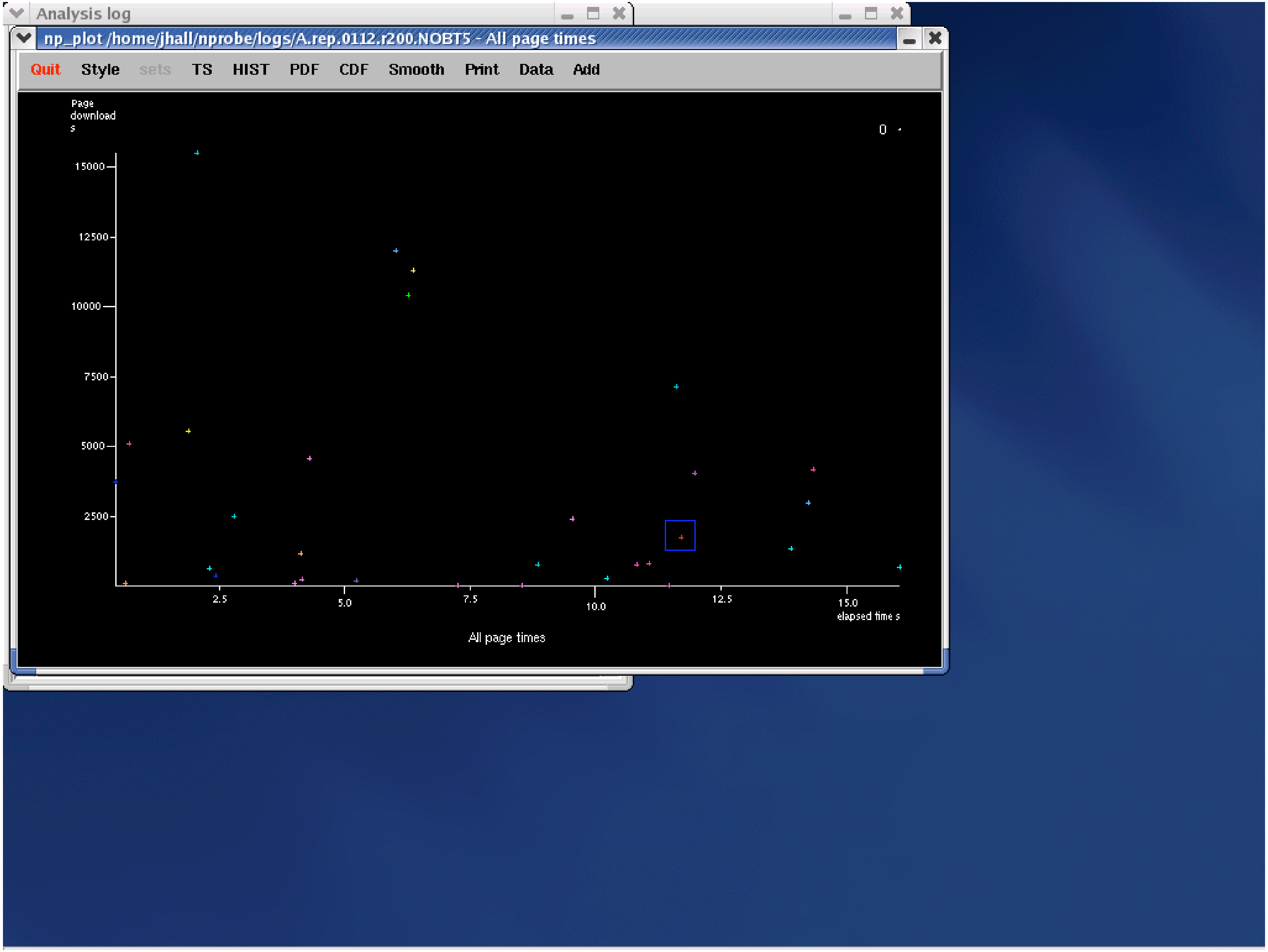
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Why?

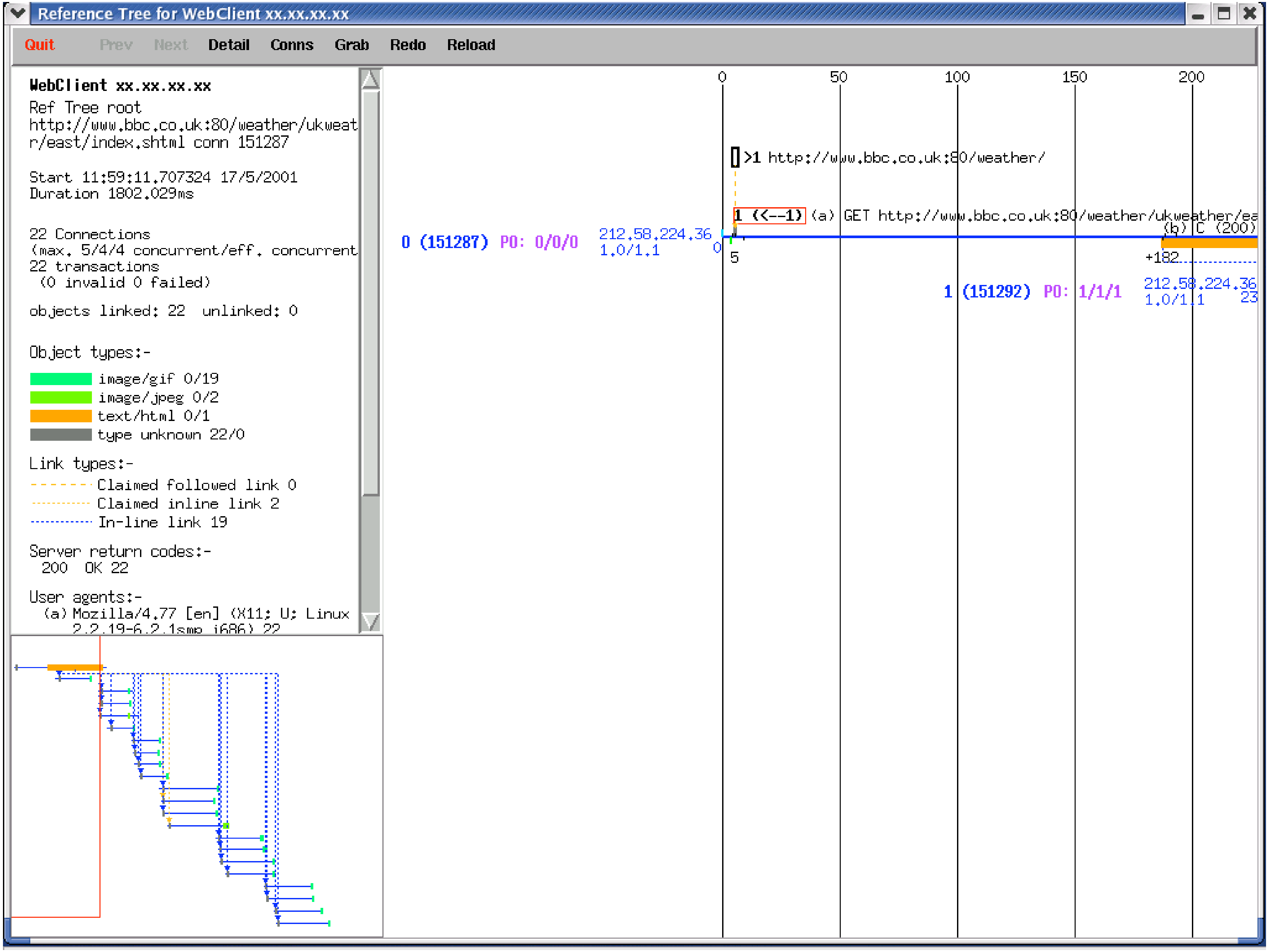
- Aids understanding
  - of observed behaviour,
  - and trends
- Teaching tool
- Debugging tool
- Maintains relationship between layers
  - tcp/ip ... http/html ... coarse statistics

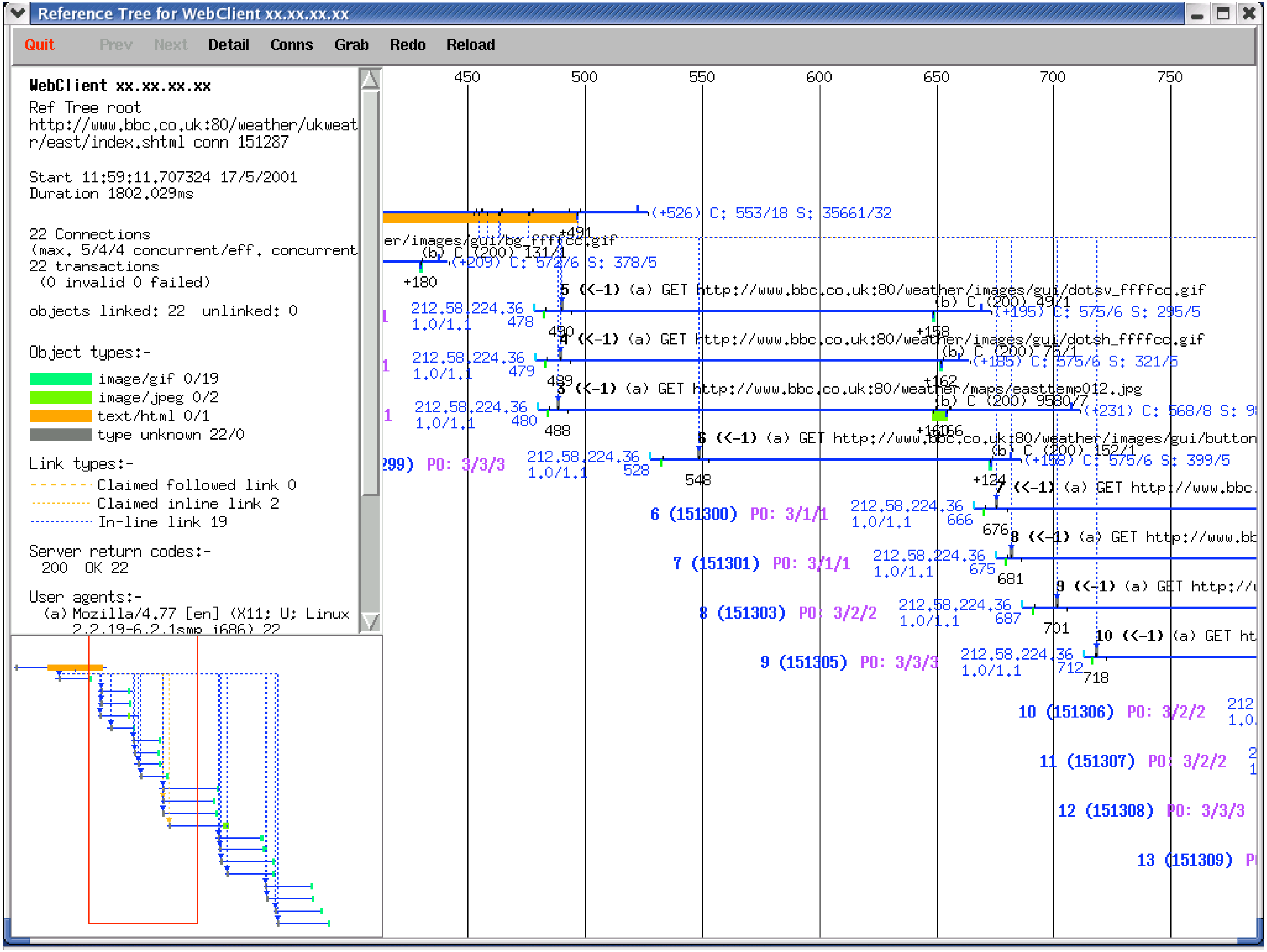

Analysis log
Quit
By_Id
By_Type
Expand
Clear
Draw
13 User_Agents seen
31 Pages seen (5 > 5 objects) 11 with delays (3 > 5 objects) 0 persistent
24 clients
8 servers
152 connections 129 objects, 0 on persistent connections 22 URLs
24 no request connections - 9 incorporated
14 objects - 28 conns 2 servers not in trees
Total connections: 199
Rejected:
Rejected by tfilter: 2
SYNs not seen 2
Total transactions 0
Valid 0
Rejected:
User Agents and inferred TCP implementations:
Servers and inferred TCP implementations:
Log saved to /home/jhall/nprobe/logs/A.rep.0112.r200.NOBT5_treelog
Ranks
Client-by-#objects
Client-by-#bytes
Client-by-#conns
Client-by-#pages
Server-by-#objects
Server-by-#bytes
Server-by-#conns
Server-by-#pages
Server-by-pdt-per-object-(>50b/P)
Server-by-pdt-per-object-(>50b/P)-less-delays
Server-by-pdt-per-object-(>50b/P)-85%
Server-by-pdt-per-object-(>50b/P)-85%-less-delays
Pages-(>5ob)-by-frequency
Pages-(>5ob)-by-download-time-per-object
Pages-(>5ob)-by-download-time-per-object-less-delays
rtmr, 0 fin-delay
tmr, 3 fin-delay
ects) 0 persistent
s 22 URLs

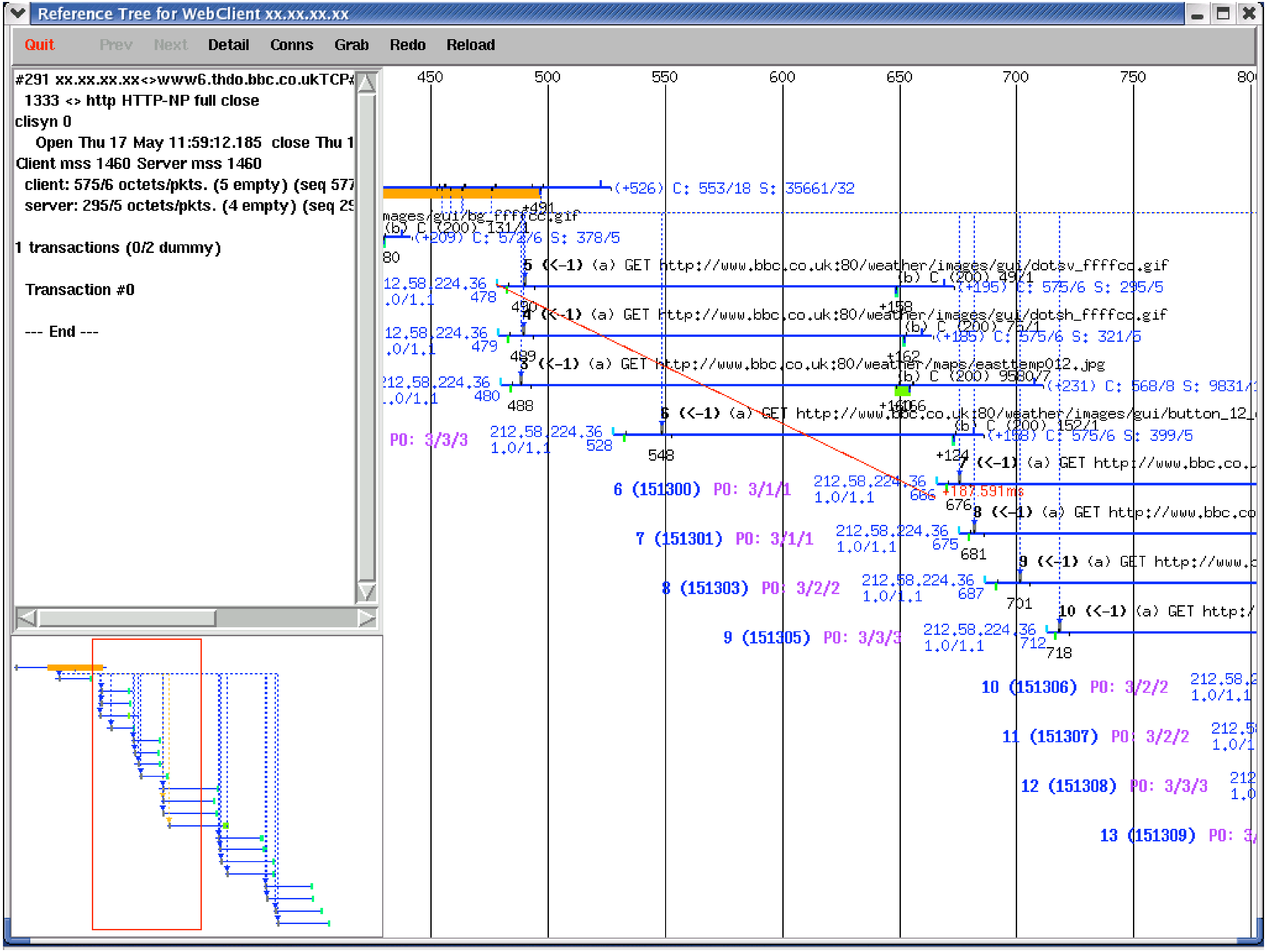

Analysis log
Quit
By_Id
By_Type
Shrink
Clear
Draw
Delays/object over time (1)
Delays/connection over time (1)
No. attempts/connection (1)
Retransmits/connection over time (1)
All Page Downloads (3)
All page downloads (31) (1)
All page downloads - pages <= 5 (26) (1)
All page downloads - pages > 5 (5) (1)
DUP SYN (20)
HIGH ACK (19)
0/1 Uconns incorporated (1)
0/4 Uconns incorporated (1)
1/2 Uconns incorporated (1)
1/3 Uconns incorporated (1)
7/14 Uconns incorporated (1)
C Retransmit (1)
Client dup syns - clients (7)
Client dup syns - conns (24)
Client dup syns - no req - clients (1)
Client dup syns - no req - conns (1)
Client dup syns - not connected - clients (2)
Client dup syns - not connected - conns (5)
Client early loss - clients (1)
Client early loss - conns (1)
Client rtmts (1)
Client transaction invalid/error (1)
DUP FIN (3)
Dummy Transaction (2)
Fail registered (1)
Large simpage first resp (4)
Link not found in referring object (3)
Connections
Page
All Pages
Page Times
85% Page Times
Page Times less delays
85% Page Times less delays
Page Times per-object
85% Page Times per-object
Page Times less delays per-object
85% Page Times less delays per-object
Retransmits
Retransmits-P-over-1s
Retransmits-P-over-5s
Retransmits-P-over-10s
Retransmits-P-over-60s
Delays
Page-delays
Page-delays-per-object
Attempts
Cancel
mr, 0 fin-delay
r, 3 fin-delay
ts) 0 persistent
22 URLs
oads - pages <= 5 (26)'

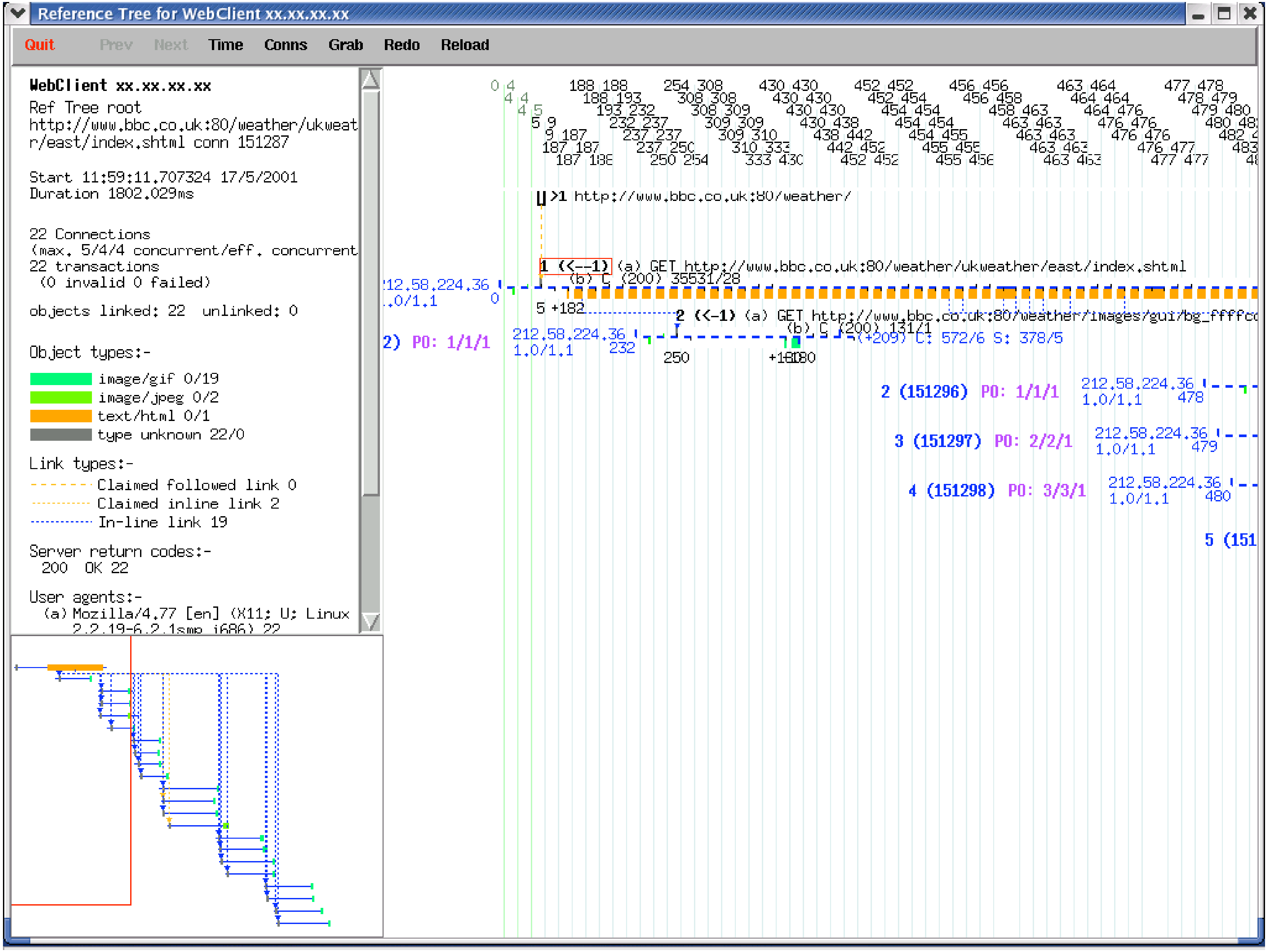


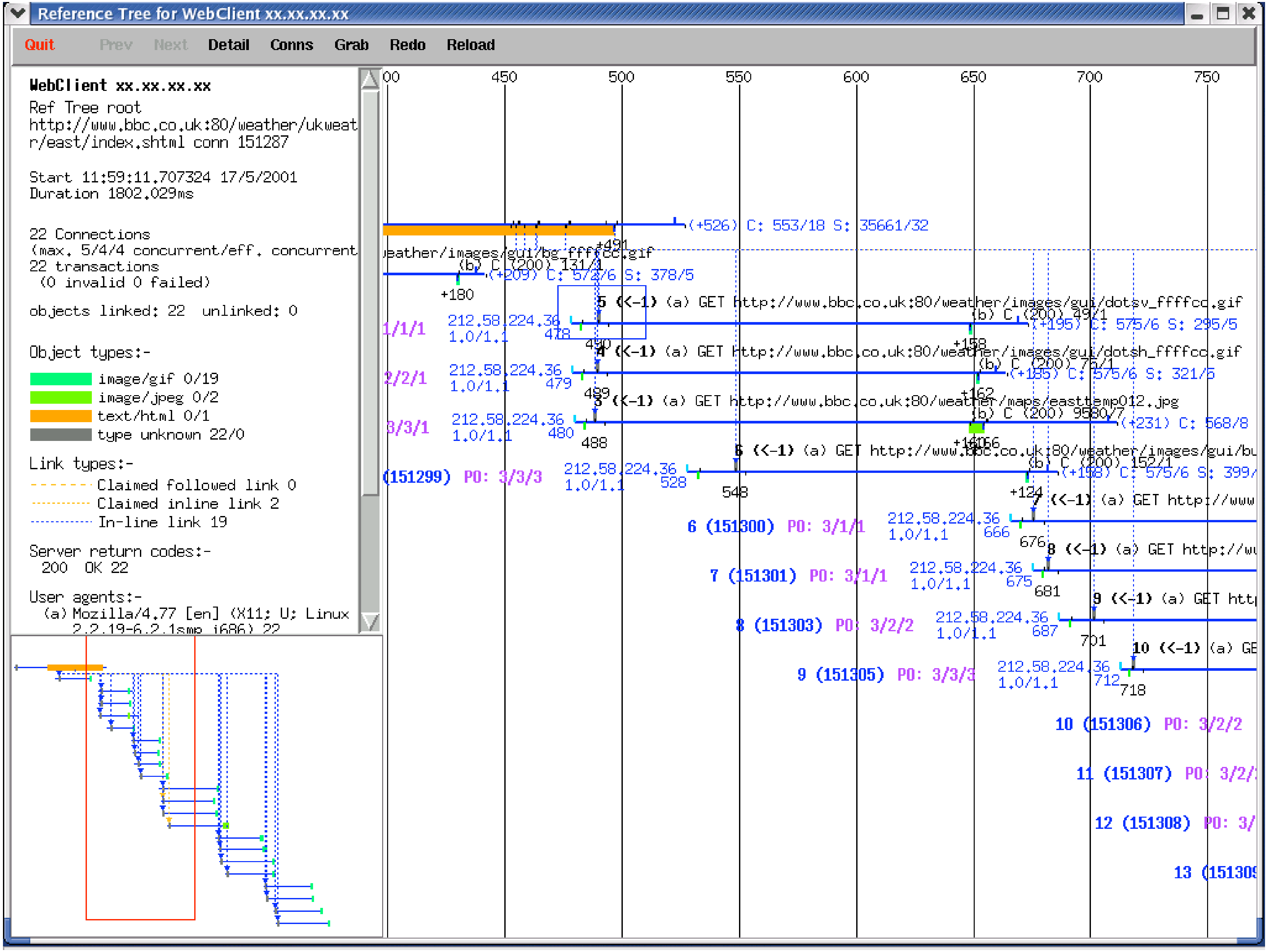
Analysis log
np_plot /home/jhall/nprobe/logs/A.rep.0112.r200.NOBT5 - All page times
Quit
Style
sets
TS
HIST
PDF
CDF
Smooth
Print
Data
Add
Page downloads
15000
12500
10000
7500
5000
2500
2.5
5.0
7.5
10.0
12.5
15.0
elapsed time s
All page times


Reference Tree for WebClient xx.xx.xx.xx
Quit
Prev
Next
Detail
Conns
Grab
Redo
Reload
WebClient xx.xx.xx.xx
Ref Tree root
http://www.bbc.co.uk:80/weather/ukweat
r/east/index.shtml conn 151287
Start 11:59:11.707324 17/5/2001
Duration 1802.029ms
22 Connections
(max. 5/4/4 concurrent/eff. concurrent
22 transactions
(0 invalid 0 failed)
objects linked: 22 unlinked: 0
Object types:-
image/gif 0/19
image/jpeg 0/2
text/html 0/1
type unknown 22/0
Link types:-
Claimed followed link 0
Claimed inline link 2
In-line link 19
Server return codes:-
200 OK 22
User agents:-
(a) Mozilla/4.77 [en] (X11; U; Linux
2.2.19-6.2.1smp i686) 22
0
50
100
150
200
>1 http://www.bbc.co.uk:80/weather/
1 (<--1) (a) GET http://www.bbc.co.uk:80/weather/ukweather/ea
(b) C (200)
0 (151287) P0: 0/0/0
212.58.224.36
1.0/1.1
0
5
+182
1 (151292) P0: 1/1/1
212.58.224.36
1.0/1.1
23

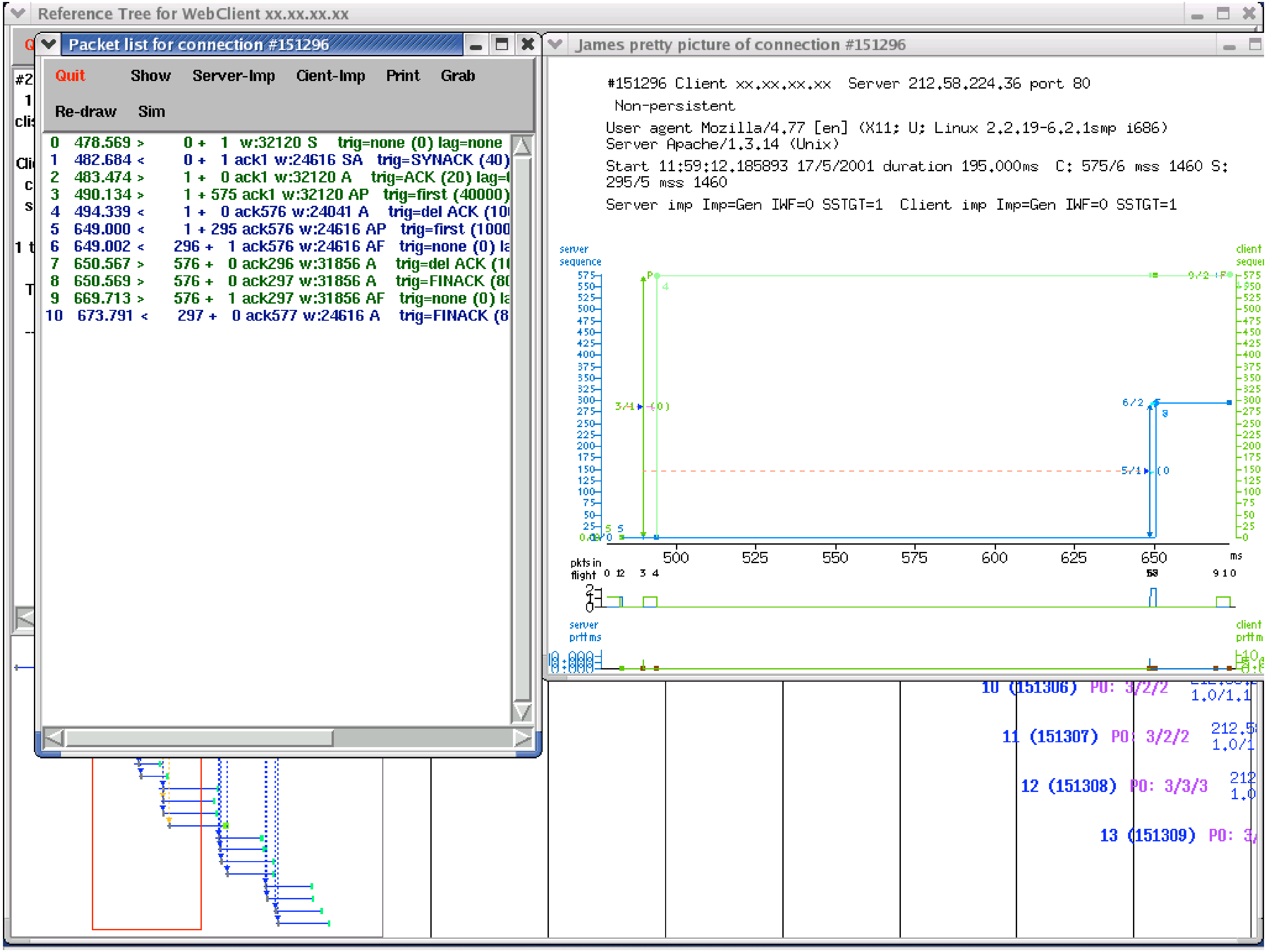

Reference Tree for WebClient xx.xx.xx.xx
Quit Prev Next Detail Conns Grab Redo Reload
WebClient xx.xx.xx.xx
Ref Tree root
http://www.bbc.co.uk:80/weather/ukweat
r/east/index.shtml conn 151287
Start 11:59:11.707324 17/5/2001
Duration 1802.029ms
22 Connections
(max. 5/4/4 concurrent/eff. concurrent
22 transactions
(0 invalid 0 failed)
objects linked: 22 unlinked: 0
Object types:-
image/gif 0/19
image/jpeg 0/2
text/html 0/1
type unknown 22/0
Link types:-
Claimed followed link 0
Claimed inline link 2
In-line link 19
Server return codes:-
200 OK 22
User agents:-
(a) Mozilla/4.77 [en] (X11; U; Linux
450 500 550 600 650 700 750
(+526) C: 553/18 S: 35661/32
5 (<-1) (a) GET http://www.bbc.co.uk:80/weather/images/gui/dotsv_ffffcc.gif
4 (<-1) (a) GET http://www.bbc.co.uk:80/weather/images/gui/dotsh_ffffcc.gif
3 (<-1) (a) GET http://www.bbc.co.uk:80/weather/maps/easttemp012.jpg
6 (151300) PO: 3/1/1
7 (151301) PO: 3/1/1
8 (151303) PO: 3/2/2
9 (151305) PO: 3/3/3
10 (151306) PO: 3/2/2
11 (151307) PO: 3/2/2
12 (151308) PO: 3/3/3
13 (151309)

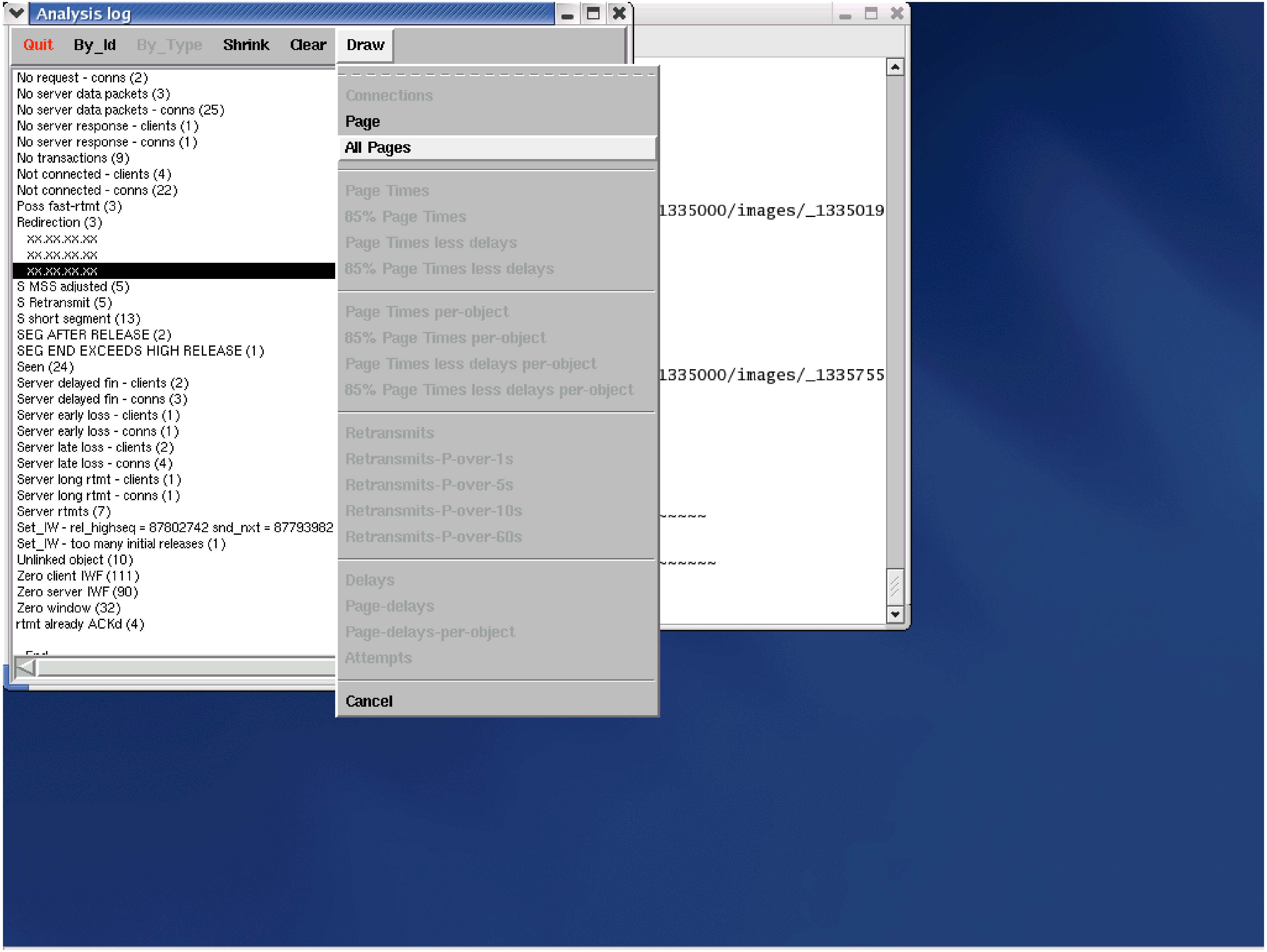

Reference Tree for WebClient xx.xx.xx.xx
Quit
Prev
Next
Detail
Conns
Grab
Redo
Reload
#291 xx.xx.xx.xx<>www6.thdo.bbc.co.ukTCP
1333 <> http HTTP-NP full close
clisyn 0
Open Thu 17 May 11:59:12.185 close Thu 1
Client mss 1460 Server mss 1460
client: 575/6 octets/pkts. (5 empty) (seq 57
server: 295/5 octets/pkts. (4 empty) (seq 2
1 transactions (0/2 dummy)
Transaction #0
--- End ---
450
500
550
600
650
700
750
5 ((-1) (a) GET http://www.bbc.co.uk:80/weather/images/gui/dotsv_ffffcc.gif
4 ((-1) (a) GET http://www.bbc.co.uk:80/weather/images/gui/dotsh_ffffcc.gif
3 ((-1) (a) GET http://www.bbc.co.uk:80/weather/maps/easttemp012.jpg
6 (151300) P0: 3/1/1
7 (151301) P0: 3/1/1
8 (151303) P0: 3/2/2
9 (151305) P0: 3/3/3
10 (151306) P0: 3/2/2
11 (151307) P0: 3/2/2
12 (151308) P0: 3/3/3
13 (151309)

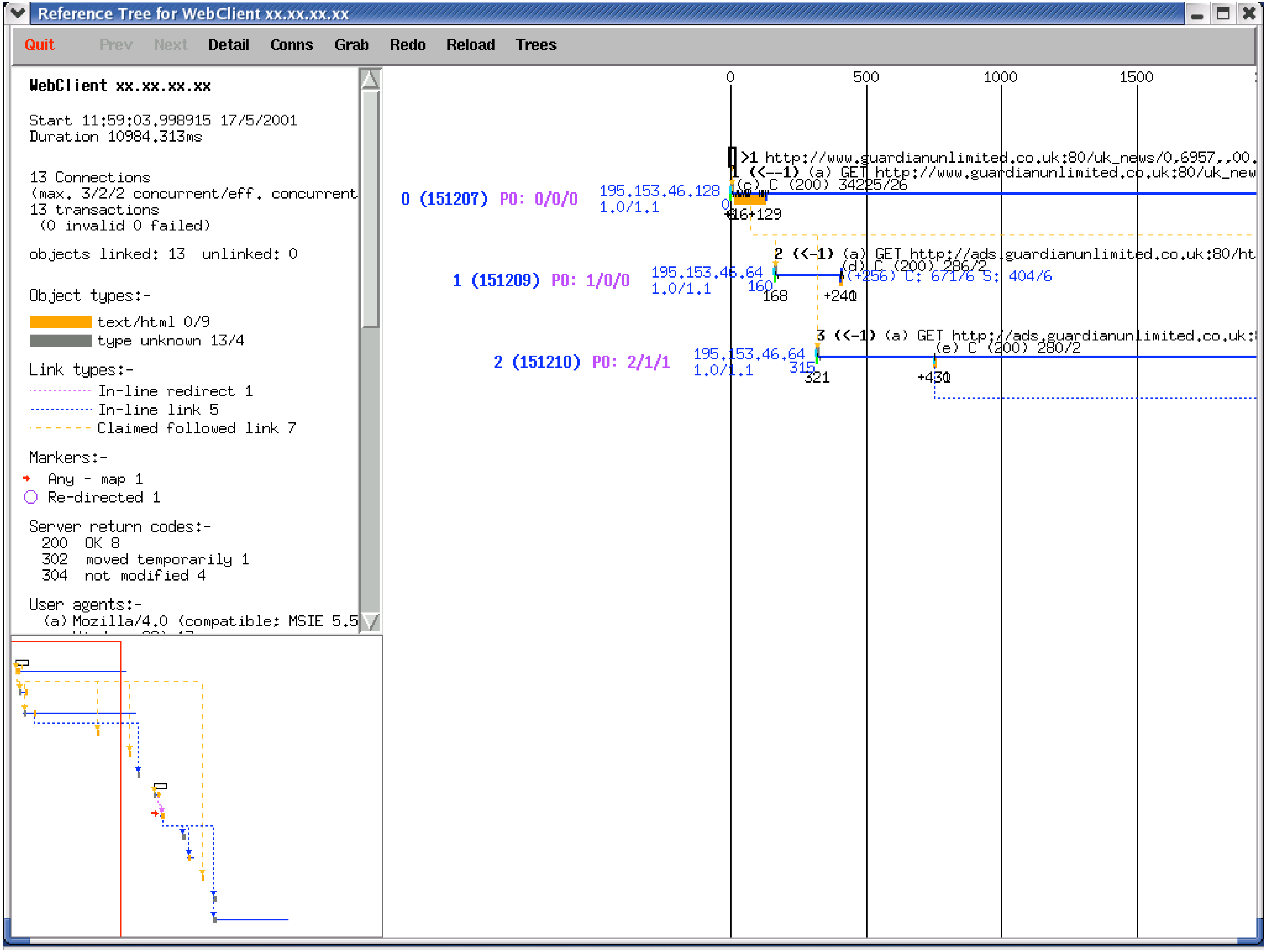

Reference Tree for WebClient xx.xx.xx.xx
Quit
Prev
Next
Time
Conns
Grab
Redo
Reload
WebClient xx.xx.xx.xx
Ref Tree root
http://www.bbc.co.uk:80/weather/ukweat
r/east/index.shtml conn 151287
Start 11:59:11.707324 17/5/2001
Duration 1802.029ms
22 Connections
(max. 5/4/4 concurrent/eff. concurrent
22 transactions
(0 invalid 0 failed)
objects linked: 22 unlinked: 0
Object types:-
image/gif 0/19
image/jpeg 0/2
text/html 0/1
type unknown 22/0
Link types:-
Claimed followed link 0
Claimed inline link 2
In-line link 19
Server return codes:-
200 OK 22
User agents:-
(a) Mozilla/4.77 [en] (X11; U; Linux
>1 http://www.bbc.co.uk:80/weather/
1 (<--1) (a) GET http://www.bbc.co.uk:80/weather/ukweather/east/index.shtml
(b) C (200) 35531/28
2 (<-1) (a) GET http://www.bbc.co.uk:80/weather/images/gui/bg_ffffcc
(b) C (200) 131/1
(+209) C: 572/6 S: 378/5
2) P0: 1/1/1
2 (151296) P0: 1/1/1
3 (151297) P0: 2/2/1
4 (151298) P0: 3/3/1
212.58.224.36
1.0/1.1

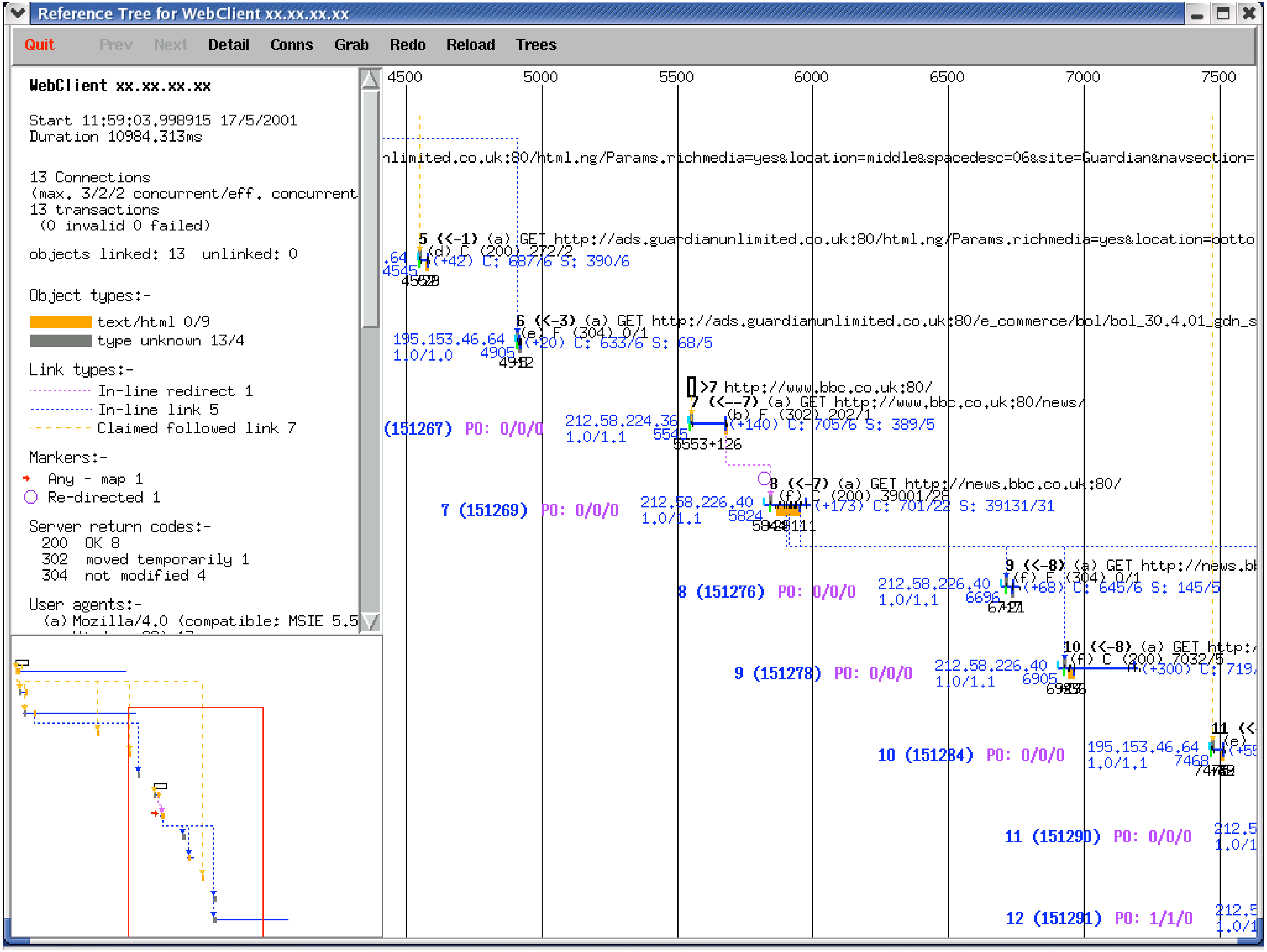


Reference Tree for WebClient xx.xx.xx.xx
Quit
Prev
Next
Detail
Conns
Grab
Redo
Reload
WebClient xx.xx.xx.xx
Ref Tree root
http://www.bbc.co.uk:80/weather/ukweat
r/east/index.shtml conn 151287
Start 11:59:11.707324 17/5/2001
Duration 1802.029ms
22 Connections
(max. 5/4/4 concurrent/eff. concurrent
22 transactions
(0 invalid 0 failed)
objects linked: 22 unlinked: 0
Object types:-
image/gif 0/19
image/jpeg 0/2
text/html 0/1
type unknown 22/0
Link types:-
Claimed followed link 0
Claimed inline link 2
In-line link 19
Server return codes:-
200 OK 22
User agents:-
(a) Mozilla/4.77 [en] (X11; U; Linux
450
500
550
600
650
700
750
(+526) C: 553/18 S: 35661/32
(+209) C: 572/6 S: 378/5
5 (<-1) (a) GET http://www.bbc.co.uk:80/weather/images/gui/dotsv_ffffcc.gif
(+195) C: 575/6 S: 295/5
4 (<-1) (a) GET http://www.bbc.co.uk:80/weather/images/gui/dotsh_ffffcc.gif
(+185) C: 575/6 S: 321/5
3 (<-1) (a) GET http://www.bbc.co.uk:80/weather/maps/easttemp012.jpg
(+231) C: 568/8
(151299) P0: 3/3/3
6 (151300) P0: 3/1/1
7 (151301) P0: 3/1/1
8 (151303) P0: 3/2/2
9 (151305) P0: 3/3/3
10 (151306) P0: 3/2/2

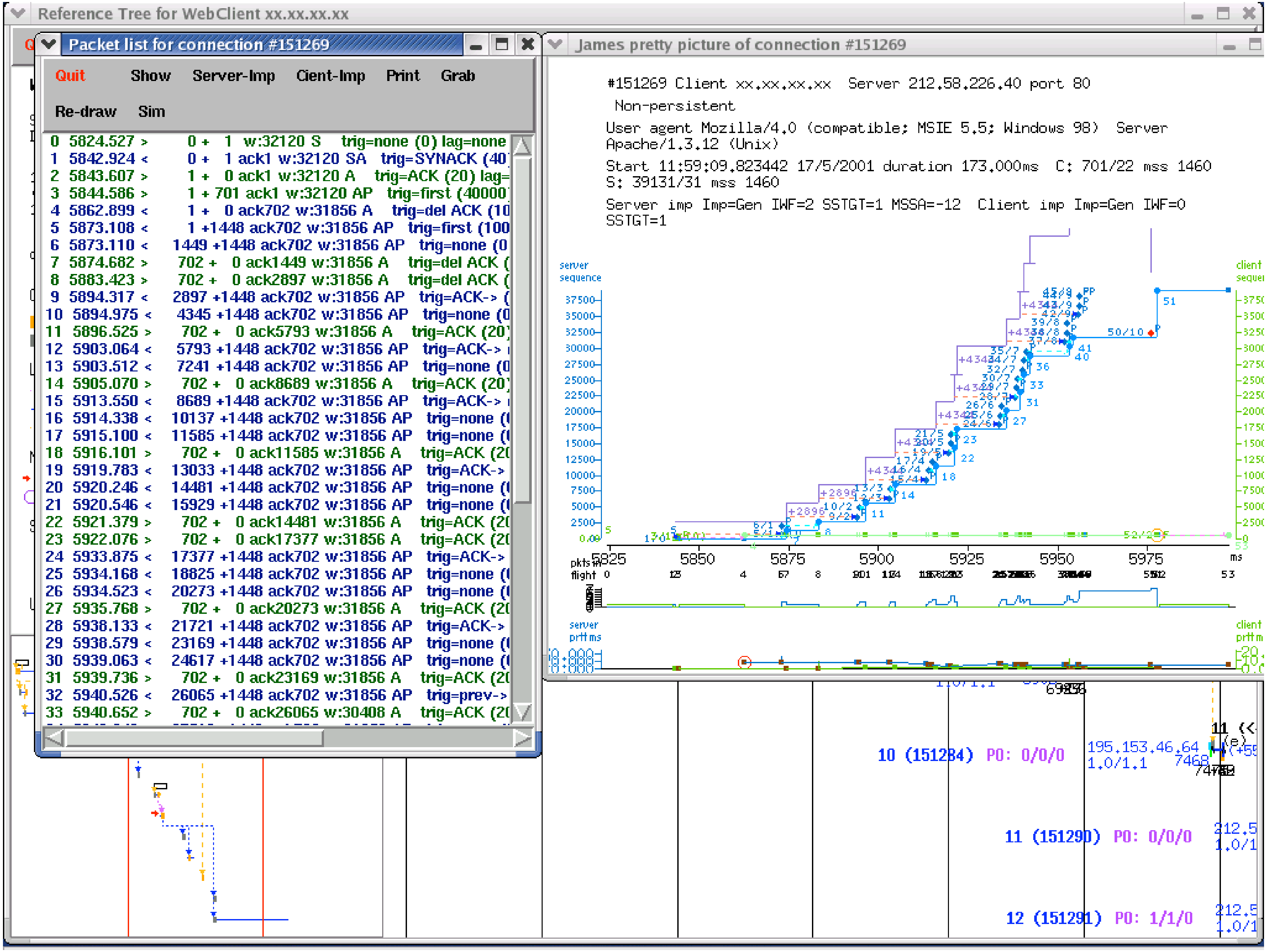

Reference Tree for WebClient xx.xx.xx.xx
Packet list for connection #151296
Quit Show Server-Imp Cient-Imp Print Grab
Re-draw Sim
0 478.569 > 0 + 1 w:32120 S trig=none (0) lag=none
1 482.684 < 0 + 1 ack1 w:24616 SA trig=SYNACK (40)
2 483.474 > 1 + 0 ack1 w:32120 A trig=ACK (20) lag=
3 490.134 > 1 + 575 ack1 w:32120 AP trig=first (40000)
4 494.339 < 1 + 0 ack576 w:24041 A trig=del ACK (10
5 649.000 < 1 + 295 ack576 w:24616 AP trig=first (1000
6 649.002 < 296 + 1 ack576 w:24616 AF trig=none (0) la
7 650.567 > 576 + 0 ack296 w:31856 A trig=del ACK (1
8 650.569 > 576 + 0 ack297 w:31856 A trig=FINACK (8
9 669.713 > 576 + 1 ack297 w:31856 AF trig=none (0) la
10 673.791 < 297 + 0 ack577 w:24616 A trig=FINACK (8
James pretty picture of connection #151296
#151296 Client xx.xx.xx.xx Server 212.58.224.36 port 80
Non-persistent
User agent Mozilla/4.77 [en] (X11; U; Linux 2.2.19-6.2.1smp i686)
Server Apache/1.3.14 (Unix)
Start 11:59:12.185893 17/5/2001 duration 195.000ms C: 575/6 mss 1460 S: 295/5 mss 1460
Server imp Imp=Gen IWF=0 SSTGT=1 Client imp Imp=Gen IWF=0 SSTGT=1
server sequence
client sequence
500 525 550 575 600 625 650 ms
pkts in flight
server prtt ms
client prtt ms
10 (151306) P0: 3/2/2
11 (151307) P0: 3/2/2
12 (151308) P0: 3/3/3
13 (151309) P0:


Analysis log
Quit
By_Id
By_Type
Shrink
Clear
Draw
No request - conns (2)
No server data packets (3)
No server data packets - conns (25)
No server response - clients (1)
No server response - conns (1)
No transactions (9)
Not connected - clients (4)
Not connected - conns (22)
Poss fast-rtmt (3)
Redirection (3)
xx.xx.xx.xx
xx.xx.xx.xx
xx.xx.xx.xx
S MSS adjusted (5)
S Retransmit (5)
S short segment (13)
SEG AFTER RELEASE (2)
SEG END EXCEEDS HIGH RELEASE (1)
Seen (24)
Server delayed fin - clients (2)
Server delayed fin - conns (3)
Server early loss - clients (1)
Server early loss - conns (1)
Server late loss - clients (2)
Server late loss - conns (4)
Server long rtmt - clients (1)
Server long rtmt - conns (1)
Server rtmts (7)
Set_IW - rel_highseq = 87802742 snd_nxt = 87793982
Set_IW - too many initial releases (1)
Unlinked object (10)
Zero client IWF (111)
Zero server IWF (90)
Zero window (32)
rtmt already ACKd (4)
Connections
Page
All Pages
Page Times
85% Page Times
Page Times less delays
85% Page Times less delays
Page Times per-object
85% Page Times per-object
Page Times less delays per-object
85% Page Times less delays per-object
Retransmits
Retransmits-P-over-1s
Retransmits-P-over-5s
Retransmits-P-over-10s
Retransmits-P-over-60s
Delays
Page-delays
Page-delays-per-object
Attempts
Cancel
1335000/images/_1335019
1335000/images/_1335755


Reference Tree for WebClient xx.xx.xx.xx
Quit
Prev
Next
Detail
Conns
Grab
Redo
Reload
Trees
WebClient xx.xx.xx.xx
Start 11:59:03.998915 17/5/2001
Duration 10984.313ms
13 Connections
(max. 3/2/2 concurrent/eff. concurrent
13 transactions
(0 invalid 0 failed)
objects linked: 13 unlinked: 0
Object types:-
text/html 0/9
type unknown 13/4
Link types:-
In-line redirect 1
In-line link 5
Claimed followed link 7
Markers:-
Any - map 1
Re-directed 1
Server return codes:-
200 OK 8
302 moved temporarily 1
304 not modified 4
User agents:-
(a) Mozilla/4.0 (compatible; MSIE 5.5
0
500
1000
1500
>1 http://www.guardianunlimited.co.uk:80/uk_news/0,6957,,00.
1 (<--1) (a) GET http://www.guardianunlimited.co.uk:80/uk_new
(c) C (200) 34225/26
0 (151207) PO: 0/0/0
195.153.46.128
1.0/1.1
+16+129
2 (<-1) (a) GET http://ads.guardianunlimited.co.uk:80/ht
(d) C (200) 286/2
1 (151209) PO: 1/0/0
195.153.46.64
1.0/1.1
160
168
(+256) C: 671/6 S: 404/6
+240
3 (<-1) (a) GET http://ads.guardianunlimited.co.uk:
(e) C (200) 280/2
2 (151210) PO: 2/1/1
195.153.46.64
1.0/1.1
315
321
+430



Reference Tree for WebClient xx.xx.xx.xx
Quit
Prev
Next
Detail
Conns
Grab
Redo
Reload
Trees
WebClient xx.xx.xx.xx
Start 11:59:03.998915 17/5/2001
Duration 10984.313ms
13 Connections
(max. 3/2/2 concurrent/eff. concurrent
13 transactions
(0 invalid 0 failed)
objects linked: 13 unlinked: 0
Object types:-
text/html 0/9
type unknown 13/4
Link types:-
In-line redirect 1
In-line link 5
Claimed followed link 7
Markers:-
Any - map 1
Re-directed 1
Server return codes:-
200 OK 8
302 moved temporarily 1
304 not modified 4
User agents:-
(a) Mozilla/4.0 (compatible; MSIE 5.5
4500
5000
5500
6000
6500
7000
7500
nlimited.co.uk:80/html.ng/Params.richmedia=yes&location=middle&spacedesc=06&site=Guardian&navsection=
5 (<-1) (a) GET http://ads.guardianunlimited.co.uk:80/html.ng/Params.richmedia=yes&location=botto
(d) C (200) 272/2
(+42) C: 687/6 S: 390/6
6 (<-3) (a) GET http://ads.guardianunlimited.co.uk:80/e_commerce/bol/bol_30.4.01_gdn_s
(e) F (304) 0/1
195.153.46.64
1.0/1.0
(+20) C: 633/6 S: 68/5
>7 http://www.bbc.co.uk:80/
7 (<--7) (a) GET http://www.bbc.co.uk:80/news/
(b) F (302) 202/1
(151267) PO: 0/0/0
212.58.224.36
1.0/1.1
(+140) C: 705/6 S: 389/5
5553+126
8 (<-7) (a) GET http://news.bbc.co.uk:80/
(f) C (200) 39001/28
7 (151269) PO: 0/0/0
212.58.226.40
1.0/1.1
(+173) C: 701/22 S: 39131/31
9 (<-8) (a) GET http://news.bb
(f) F (304) 0/1
8 (151276) PO: 0/0/0
212.58.226.40
1.0/1.1
(+68) C: 645/6 S: 145/5
10 (<-8) (a) GET http:/
(f) C (200) 7032/5
9 (151278) PO: 0/0/0
212.58.226.40
1.0/1.1
(+300) C: 719/
11 (<
(e)
10 (151284) PO: 0/0/0
195.153.46.64
1.0/1.1
11 (151290) PO: 0/0/0
12 (151291) PO: 1/1/0

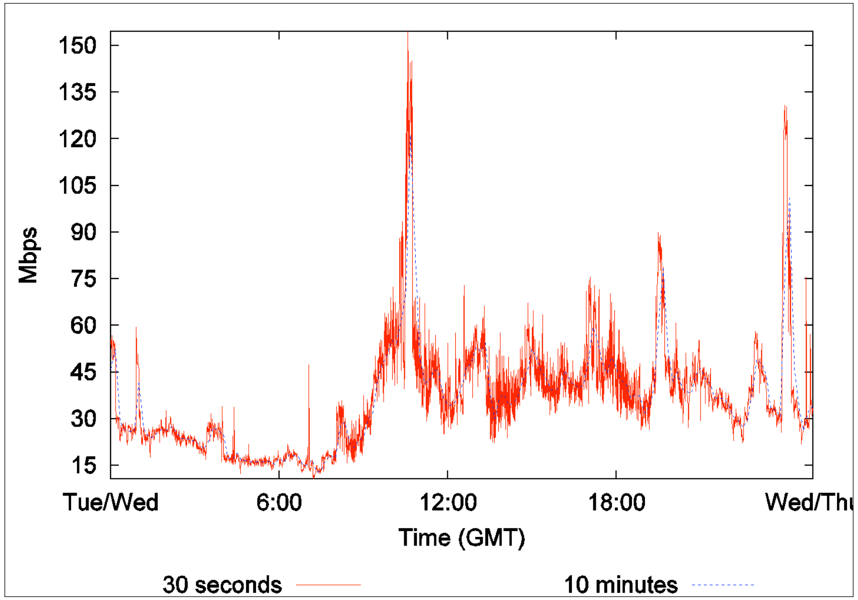

Reference Tree for WebClient xx.xx.xx.xx
Packet list for connection #151269
Quit Show Server-Imp Cient-Imp Print Grab
Re-draw Sim
0 5824.527 > 0 + 1 w:32120 S trig=none (0) lag=none
1 5842.924 < 0 + 1 ack1 w:32120 SA trig=SYNACK (40
2 5843.607 > 1 + 0 ack1 w:32120 A trig=ACK (20) lag=
3 5844.586 > 1 + 701 ack1 w:32120 AP trig=first (40000
4 5862.899 < 1 + 0 ack702 w:31856 A trig=del ACK (10
5 5873.108 < 1 +1448 ack702 w:31856 AP trig=first (100
6 5873.110 < 1449 +1448 ack702 w:31856 AP trig=none (0
7 5874.682 > 702 + 0 ack1449 w:31856 A trig=del ACK (
8 5883.423 > 702 + 0 ack2897 w:31856 A trig=del ACK (
9 5894.317 < 2897 +1448 ack702 w:31856 AP trig=ACK-> (
10 5894.975 < 4345 +1448 ack702 w:31856 AP trig=none (0
11 5896.525 > 702 + 0 ack5793 w:31856 A trig=ACK (20)
12 5903.064 < 5793 +1448 ack702 w:31856 AP trig=ACK->
13 5903.512 < 7241 +1448 ack702 w:31856 AP trig=none (0
14 5905.070 > 702 + 0 ack8689 w:31856 A trig=ACK (20)
15 5913.550 < 8689 +1448 ack702 w:31856 AP trig=ACK->
16 5914.338 < 10137 +1448 ack702 w:31856 AP trig=none (
17 5915.100 < 11585 +1448 ack702 w:31856 AP trig=none (
18 5916.101 > 702 + 0 ack11585 w:31856 A trig=ACK (2
19 5919.783 < 13033 +1448 ack702 w:31856 AP trig=ACK->
20 5920.246 < 14481 +1448 ack702 w:31856 AP trig=none (
21 5920.546 < 15929 +1448 ack702 w:31856 AP trig=none (
22 5921.379 > 702 + 0 ack14481 w:31856 A trig=ACK (2
23 5922.076 > 702 + 0 ack17377 w:31856 A trig=ACK (2
24 5933.875 < 17377 +1448 ack702 w:31856 AP trig=ACK->
25 5934.168 < 18825 +1448 ack702 w:31856 AP trig=none (
26 5934.523 < 20273 +1448 ack702 w:31856 AP trig=none (
27 5935.768 > 702 + 0 ack20273 w:31856 A trig=ACK (2
28 5938.133 < 21721 +1448 ack702 w:31856 AP trig=ACK->
29 5938.579 < 23169 +1448 ack702 w:31856 AP trig=none (
30 5939.063 < 24617 +1448 ack702 w:31856 AP trig=none (
31 5939.736 > 702 + 0 ack23169 w:31856 A trig=ACK (2
32 5940.526 < 26065 +1448 ack702 w:31856 AP trig=prev->
33 5940.652 > 702 + 0 ack26065 w:30408 A trig=ACK (2
James pretty picture of connection #151269
#151269 Client xx.xx.xx.xx Server 212.58.226.40 port 80
Non-persistent
User agent Mozilla/4.0 (compatible; MSIE 5.5; Windows 98) Server Apache/1.3.12 (Unix)
Start 11:59:09.823442 17/5/2001 duration 173.000ms C: 701/22 mss 1460 S: 39131/31 mss 1460
Server imp Imp=Gen IWF=2 SSTGT=1 MSSA=-12 Client imp Imp=Gen IWF=0 SSTGT=1
server sequence
client sequence
5825 5850 5875 5900 5925 5950 5975 ms
pkts in flight
server prtt ms
client prtt ms
10 (151284) P0: 0/0/0
195.153.46.64 1.0/1.1
11 (151290) P0: 0/0/0
12 (151291) P0: 1/1/0

Example 2: Experiences at a GRID Access Point

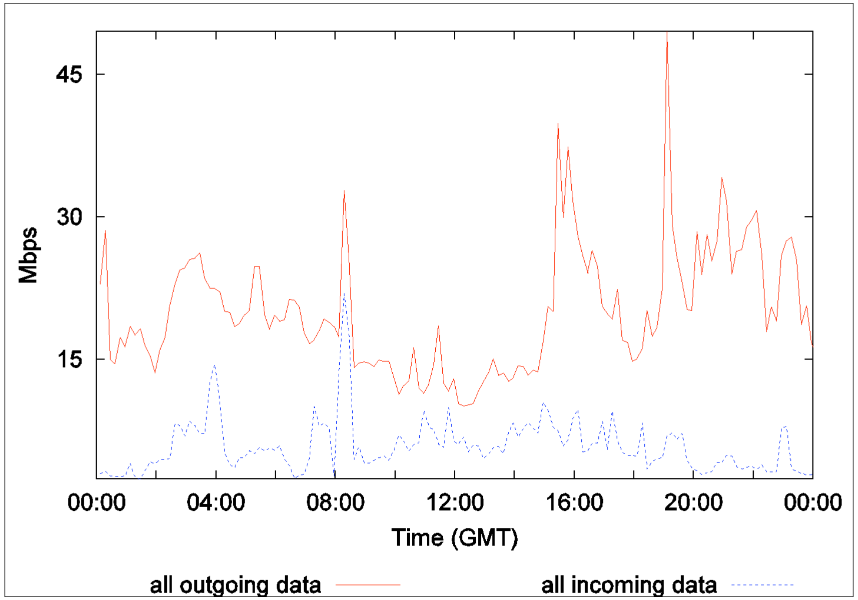
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

- Research community of 1,000 on their own campus
- Significant (unique) data provided by this site to the world community
- One of three sites where data is continuously updated – so data is continuously transferred between partners and downloaded by collaborators

Traffic to/from access point



Total Link traffic



Each Direction

Contrasting port and content based classification

	Port-based		Content-based	
	Packets	Bytes	Packets	Bytes
BULK	49.97	45.00	65.06	64.94
DATABASE	0.03	0.03	0.84	0.96
GRID	0.03	0.07	0.00	0.00
INTERACTIVE	1.19	0.43	0.75	0.39
MAIL	3.37	3.62	3.37	3.62
SERVICES	0.07	0.02	0.29	0.28
WWW	19.98	20.40	26.50	27.60
UNKNOWN	28.36	30.44	-	-
OTHER	-	-	3.20	3.11

Overheads vs. Accuracy

From Moore, Papagiannaki submission to IMC

(measures in packets)

Method	UNKNOWN	Correctly Identified
Port	29%	71%
1KB Signature	24%	74%
1KB Protocol Parse	19%	81%
Important flow Assembly/Parse	1%	98%
Full Assemble/Parse	0%	100%

Classification Surprises

- Significant asymmetry to/from site
- Port-based classification was **so** wrong
- Considered the most important node for it's work yet,
- No GRID-application traffic!

  It was all GRID web services or FTP traffic

(For the GRID community this was surprising, for the rest of us – less so.)

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Conclusions

- Our approach is sound: the implementation works and has been perfectly satisfactory for the environments into which we have deployed
- Clear avenues of development are available to us
- Further work with deployments will provide input to this work and provide data for other projects too

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

What have we learnt?

- Always things to improve:
  - hardware
  - optimization

- Important to remember:
  - compression of between 1:12 and 1:50 is achieved
  - output data is all ready for off-line processing

Next...

- Continue work to identify and quantify GRID and Internet applications:

  e.g., Access GRID

- Evaluating 10 Gbps scheme
  - using test environments
  - considering deployment (10Gbps surprisingly uncommon)

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Enabling...

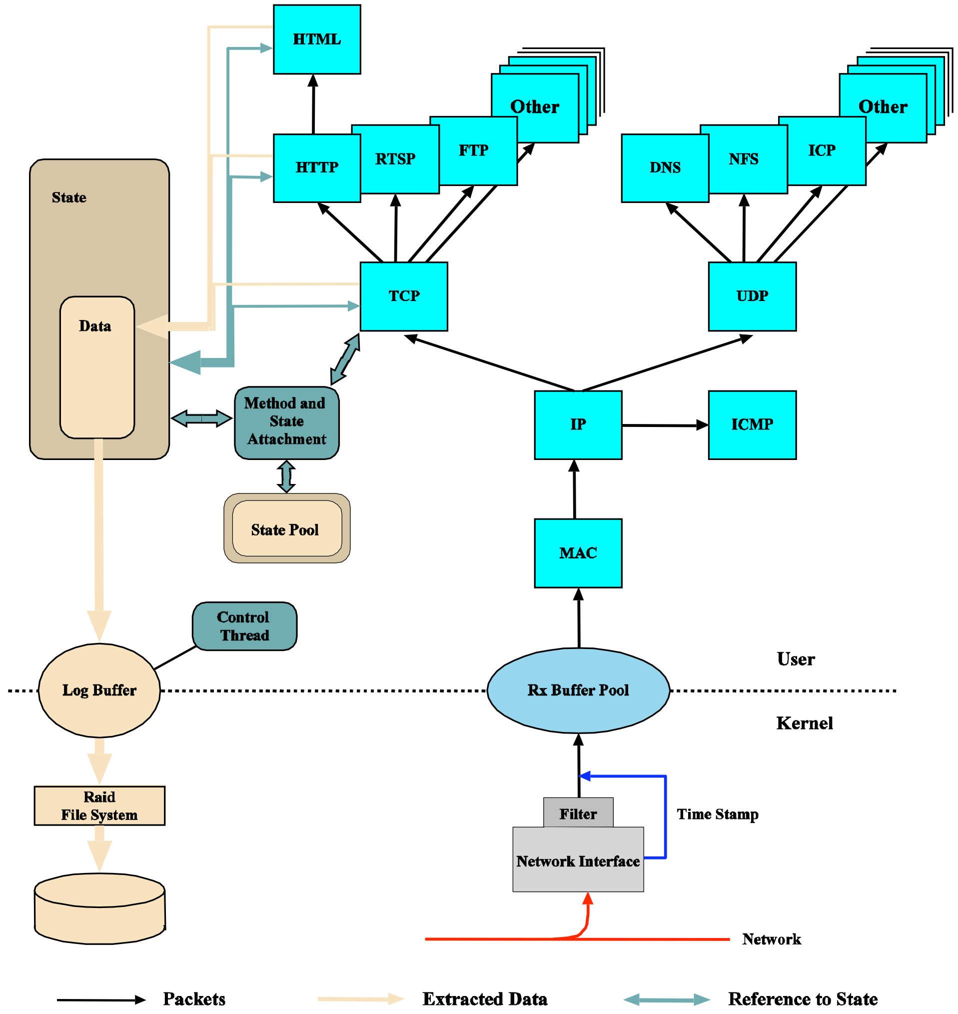
- Using of 10Gbps for the new UKLIGHT testbed – following the growth from implementation to deployment into full use.

- Current work enables us to assess suitability of Peer-2-peer algorithms for distributing data currently shared using FTP...

Backup slides

Architecture

- Data driven model
- Single-thread model to maximize efficiency
- Avoid memory copy when practical



Monitor / Hardware

Objective: use commodity hardware even the NIC

- modify the firmware rather than building hardware
- add/use time-stamping – currently 1_s
- perform filtering on card with minimal overhead
- Current 1Gbps cards supported:
  - Alteon / 3Com 3c985B
  - SysKonnect sk98xx

Monitor / Hardware Filtering

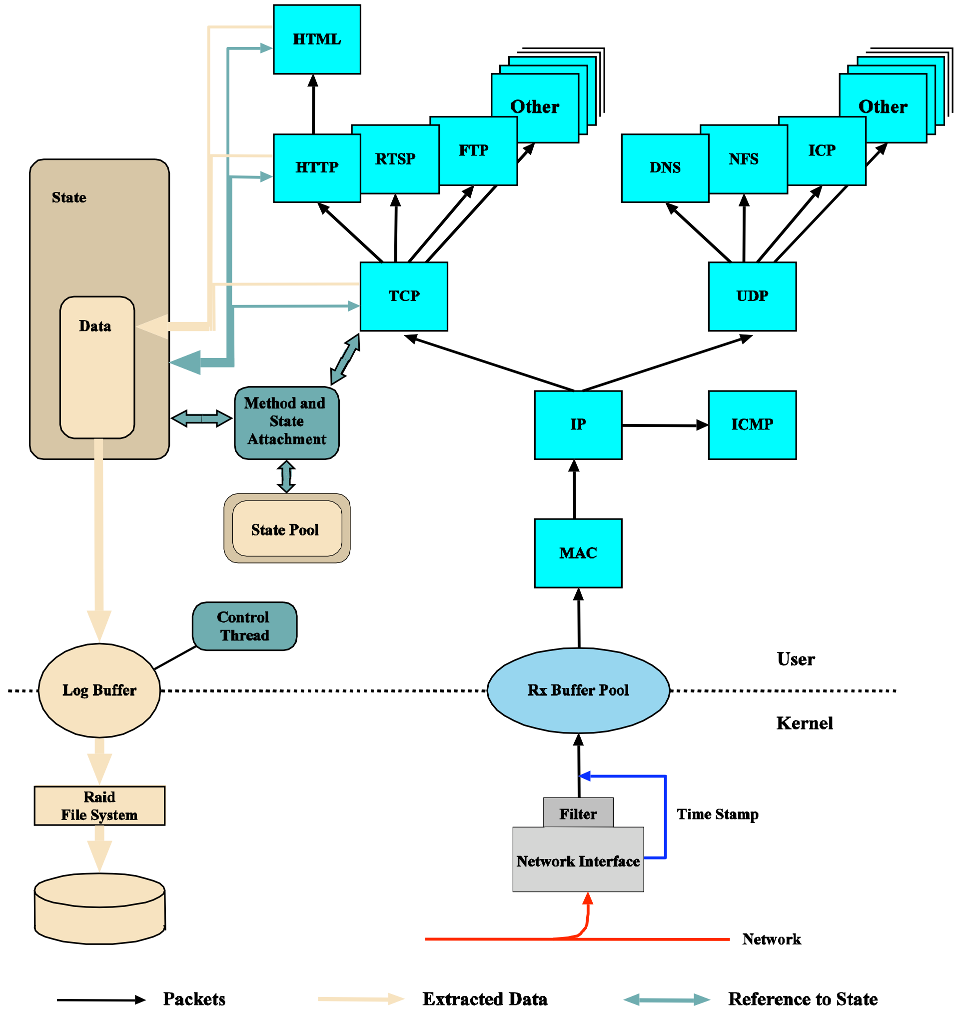
- Using hash of XOR of SRC/DEST as a selection criterion:
  - our approach works best when both directions of traffic are kept together.
- Work in progress
  - how often do we need to update filters?
  - what can we optimize filters for:
    - filter size?
    - packet distribution?
    - equalizing flows? packets? bytes?
    - This problem is common to the load-sharing community

Receive FIFO Implementation

Receive Buffer is not a FIFO

- This means that data-processing mechanisms can return data buffers when they are finished
- Out-of-order return allows easier handling of packet loss and packet reordering

- Discards packets when memory runs low
- Implemented to hang on to packets in case of potential use or reuse



Monitor / Processing

Compress/discard where we can:

- network, TCP and application layers can each have considerable temporal redundancy
- application-specific reductions such as removing the data object from http transactions – we keep a fingerprint of the object so as we can recognise the same object even with different URLs
- loss-less compression (lz, gzip, etc.)

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Nprobe modules

- Current
  - TCP and UDP on IP
  - HTTP and HTML
  - DNS
  - FTP
- Past (deprecated)
  - NFS (v2)